(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number
WO 01/72761 A2

(51) International Patent Classification[7]: C07H

(21) International Application Number: PCT/US01/08283

(22) International Filing Date: 15 March 2001 (15.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/192,127 24 March 2000 (24.03.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHEW, Anne** [US/US]; 1477 Beacon Street, #64, Brookline, MA 02446 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US).

(74) Agents: **SHANER, Sandra, L.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/72761 A2

(54) Title: HAPLOTYPES OF THE PLTP GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human phospholipid transfer protein (PLTP) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the PLTP gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the PLTP gene in an individual are also disclosed. Polynucleotides containing one or more of the PLTP polymorphisms disclosed herein are also described.

HAPLOTYPES OF THE PLTP GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/192,127 filed March 24, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human phospholipid transfer protein (PLTP) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical

trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* 15:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* 1:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* 4:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* 5:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the primary sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of atherosclerosis is the phospholipid transfer protein (PLTP) gene or its encoded product. PLTP is an important regulator of high-density lipoprotein (HDL) metabolism. The two main functions of PLTP are the transfer of phospholipids between lipoprotein particles and the modulation of HDL size and composition (a process called HDL conversion) (Huuskonen et al., *Biochemistry* 2000; 39:16092-16098). HDL is responsible for reverse cholesterol transport, which is the process of transporting cellular cholesterol back to the liver where the

excess is converted into bile acids, keeping the plasma cholesterol levels at a minimum.

In a PLTP-overexpressing mouse model, a 2-3 fold increase was seen in the potential to form pre β-HDL, which acts as an initial acceptor of cellular cholesterol from the plasma membrane, as compared to wild-type mice (Huuskonen and Ehnholm, *Curr. Opin. Lipidol.* 2000; 11:285-289). In addition, plasma from the transgenic mice was much more efficient in preventing the accumulation of intracellular cholesterol in macrophages than plasma from wild-type mice.

Other studies have shown that homozygous PLTP knock-out mice had a marked decrease in HDL phospholipids, cholesterol, and apolipoprotein A1 (Huuskonen and Ehnholm, *Curr. Opin. Lipidol.* 2000; 11:285-289). Several studies have reported that a decrease in HDL results in an accumulation of plasma cholesterol leading to atherosclerosis (Stein and Stein, 1999, *Atherosclerosis* 144:285-301). On a high fat diet, these mice also showed an increase in low-density lipoprotein (LDL) phospholipids, plasma cholesterol and cholesteryl ester (Huuskonen and Ehnholm, *Curr. Opin. Lipidol.* 2000; 11:285-289). Taken together, these results suggest that PLTP modulates the antiatherogenic effects of HDL.

The phospholipid transfer protein gene is located on chromosome 20 and contains 16 exons that encode a 493 amino acid protein. A reference sequence for the PLTP gene is shown in Figure 1 (GenBank Accession No. AL008726.1; SEQ ID NO:1). Reference sequences for the coding sequence (GenBankAccession No. NM_006227.1) and protein are shown in Figures 2 (SEQ ID NO:2) and 3 (SEQ ID NO:3), respectively.

Because of the potential for variation in the PLTP gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the PLTP gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of PLTP as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 25 novel polymorphic sites in the PLTP gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the indicated GenBank Accession Number: 1128 (PS1), 1632 (PS2), 1760 (PS3), 2047 (PS4), 2146 (PS5), 2279 (PS6), 2634 (PS7), 3840 (PS8), 3884 (PS9), 3897 (PS10), 4055 (PS11), 4308 (PS12), 7462 (PS13), 8710 (PS14), 8955 (PS15), 10907 (PS16), 11040 (PS17), 11058 (PS18), 11424 (PS19), 11486 (PS20), 13913 (PS21), 13936 (PS22), 13982 (PS23), 14559 (PS24) and 14714 (PS25) in AL008726.1. The polymorphisms at these sites are guanine or adenine at PS1, guanine or cytosine at PS2, guanine or cytosine at PS3, cytosine or thymine at PS4, thymine or cytosine at PS5, adenine or guanine at PS6, adenine or guanine at PS7, cytosine or adenine at PS8, cytosine or thymine at PS9, adenine or guanine at PS10, cytosine or adenine at PS11, cytosine or thymine at PS12, adenine or guanine at PS13, guanine or adenine at PS14, guanine or thymine at PS15, adenine or cytosine at PS16, adenine or guanine at PS17, cytosine or thymine at PS18, cytosine or guanine at PS19, cytosine or adenine at PS20, cytosine or guanine at PS21,

guanine or thymine at PS22, cytosine or thymine at PS23, guanine or adenine at PS24 and thymine or cytosine at PS25. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS25 in the PLTP gene, which are shown below in Tables 5 and 4, respectively. Each of these PLTP haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the PLTP gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the PLTP gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25 in both copies of the PLTP gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel PLTP polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel PLTP polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the PLTP gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the PLTP gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's PLTP gene is defined by one of the PLTP haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's PLTP gene are defined by one of the PLTP haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the PLTP haplotype or haplotype pair of an individual is useful for improving the efficiency and outcome of several steps in the discovery and development of drugs for treating diseases associated with PLTP activity, e.g., atherosclerosis.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate PLTP as a candidate target for treating a specific condition or disease predicted to be associated with PLTP activity. Determining for a particular population the frequency of one or more of the individual PLTP haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue PLTP as a target for treating the specific disease of interest. In particular, if variable PLTP activity is associated with the disease, then one or more PLTP haplotypes or haplotype pairs will be

found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed PLTP haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable PLTP activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any PLTP haplotype or haplotype pair, apply the information derived from detecting PLTP haplotypes in an individual to decide whether modulating PLTP activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting PLTP to treat a specific condition or disease predicted to be associated with PLTP activity. For example, detecting which of the PLTP haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent PLTP isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular PLTP haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the PLTP gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with PLTP activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the PLTP haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute PLTP haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a PLTP haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any PLTP haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a PLTP genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the PLTP genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the PLTP genotype, haplotype, or haplotype pair in a reference population. A higher frequency of the PLTP genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the PLTP genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the PLTP haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and

therapeutic treatments for atherosclerosis.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the PLTP gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, cytosine at PS5, guanine at PS6, guanine at PS7, adenine at PS8, thymine at PS9, guanine at PS10, adenine at PS11, thymine at PS12, guanine at PS13, adenine at PS14, thymine at PS15, cytosine at PS16, guanine at PS17, thymine at PS18, guanine at PS19, adenine at PS20, guanine at PS21, thymine at PS22, thymine at PS23, adenine at PS24 and cytosine at PS25.

A particularly preferred polymorphic variant is an isogene of the PLTP gene. A PLTP isogene of the invention comprises guanine or adenine at PS1, guanine or cytosine at PS2, guanine or cytosine at PS3, cytosine or thymine at PS4, thymine or cytosine at PS5, adenine or guanine at PS6, adenine or guanine at PS7, cytosine or adenine at PS8, cytosine or thymine at PS9, adenine or guanine at PS10, cytosine or adenine at PS11, cytosine or thymine at PS12, adenine or guanine at PS13, guanine or adenine at PS14, guanine or thymine at PS15, adenine or cytosine at PS16, adenine or guanine at PS17, cytosine or thymine at PS18, cytosine or guanine at PS19, cytosine or adenine at PS20, cytosine or guanine at PS21, guanine or thymine at PS22, cytosine or thymine at PS23, guanine or adenine at PS24 and thymine or cytosine at PS25. The invention also provides a collection of PLTP isogenes, referred to herein as a PLTP genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a PLTP cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 47, adenine at a position corresponding to nucleotide 371, adenine at a position corresponding to nucleotide 849 and cytosine at a position corresponding to nucleotide 1452. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a PLTP isogene defined by haplotypes 1- 23 and 25.

Polynucleotides complementary to these PLTP genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the PLTPgene will be useful in studying the expression and function of PLTP, and in expressing PLTP protein for use in screening for candidate drugs to treat diseases related to PLTP activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express PLTP for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the PLTP protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid

selected from the group consisting of arginine at a position corresponding to amino acid position 16 and tyrosine at a position corresponding to amino acid position 124. A polymorphic variant of PLTP is useful in studying the effect of the variation on the biological activity of PLTP as well as on the binding affinity of candidate drugs targeting PLTP for the treatment of atherosclerosis.

The present invention also provides antibodies that recognize and bind to the above polymorphic PLTP protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the PLTP polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the PLTP isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against PLTP protein, and for testing the efficacy of therapeutic agents and compounds for atherosclerosis in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the PLTP gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the PLTP gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing PLTP haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the PLTP gene (Genbank Accession Number AL008726.1; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:129 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R = G or A, Y = T or C, M = A or C, K = G or T, S = G or C, and W = A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the PLTP coding sequence (contiguous lines; SEQ ID NO:2) with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the PLTP protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the PLTP gene. As

described in more detail below, the inventors herein discovered 25 isogenes of the PLTP gene by characterizing the PLTP gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The PLTP isogenes present in the human reference population are defined by haplotypes for 25 polymorphic sites in the PLTP gene, all of which are believed to be novel. The PLTP polymorphic sites identified by the inventors are referred to as PS1-PS25 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25. Using the genotypes identified in the Index Repository for PS1-PS25 and

the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the PLTP gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the PLTP gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether PLTP is a suitable target for drugs to treat atherosclerosis, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the PLTP gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel PLTP polymorphisms and haplotypes identified herein.

The compositions comprise at least one PLTP genotyping oligonucleotide. In one embodiment, a PLTP genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a PLTP polynucleotide, i.e., a PLTP isogene. As used herein, specific hybridization

means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-PLTP polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the PLTP gene using the polymorphism information provided herein in conjunction with the known sequence information for the PLTP gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th

position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting PLTP gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.: AL008726.1

TGTGCCTRGACCTGG (SEQ ID NO:4) and its complement,
CTATCCCSTTCCCAG (SEQ ID NO:5) and its complement,
TGAGGGASCCGGGAG (SEQ ID NO:6) and its complement,
CGCATCGYCTCTCTG (SEQ ID NO:7) and its complement,
GGCTAGTYCCCACTT (SEQ ID NO:8) and its complement,
GGCGCACRTGCAGAG (SEQ ID NO:9) and its complement,
GAGTGGGRTGGGGCG (SEQ ID NO:10) and its complement,
GAGTTTCMGTTTCTT (SEQ ID NO:11) and its complement,
TTTCATAYGAGTTGA (SEQ ID NO:12) and its complement,
GATAGGARGGTTTGG (SEQ ID NO:13) and its complement,
GGTGTGTMCATCCGC (SEQ ID NO:14) and its complement,
GTGTGAAYGTGGCCA (SEQ ID NO:15) and its complement,
TAACAAGRCCTCTTT (SEQ ID NO:16) and its complement,
TCCGGGCRGGGGCCC (SEQ ID NO:17) and its complement,
AGGTGGGKCCCCCAG (SEQ ID NO:18) and its complement,
GTGGCAAMGGCATGA (SEQ ID NO:19) and its complement,
CAAAGGARATGGCAG (SEQ ID NO:20) and its complement,
CTTGACCYGCCCCTG (SEQ ID NO:21) and its complement,
AGGTAGGSAGGCGCA (SEQ ID NO:22) and its complement,
GATGTTGMGCTTAAA (SEQ ID NO:23) and its complement,
ATCCCCCSCTCCTCA (SEQ ID NO:24) and its complement,
CTCACCTKCGGGCAG (SEQ ID NO:25) and its complement,
CTGTTTCYGCCCTCG (SEQ ID NO:26) and its complement,
GTCTCTTRATCTCTA (SEQ ID NO:27) and its complement, and
CGTCCACYGCCCCCA (SEQ ID NO:28) and its complement.

A preferred ASO primer for detecting PLTP gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AL008726.1

TGACTTTGTGCCTRG (SEQ ID NO:29); TTACAACCAGGTCYA (SEQ ID NO:30);
GCGTCCCTATCCCST (SEQ ID NO:31); GAGGGACTGGGAASG (SEQ ID NO:32);
AGGACCTGAGGGASC (SEQ ID NO:33); CCGGGTCTCCCGGST (SEQ ID NO:34);
GCAAGTCGCATCGYC (SEQ ID NO:35); GAGGCTCAGAGAGRC (SEQ ID NO:36);
GATCCAGGCTAGTYC (SEQ ID NO:37); AACCATAAGTGGGRA (SEQ ID NO:38);
CTGGCAGGCGCACRT (SEQ ID NO:39); TGGGAACTCTGCAYG (SEQ ID NO:40);

GAGGCTGAGTGGGRT (SEQ ID NO:41); TGGCCACGCCCCAYC (SEQ ID NO:42);
CTTCCTGAGTTTCMG (SEQ ID NO:43); CAGATGAAGAAACKG (SEQ ID NO:44);
TATGCATTTCATAYG (SEQ ID NO:45); TTCCTATCAACTCRT (SEQ ID NO:46);
CGAGTTGATAGGARG (SEQ ID NO:47); ATCCTACCAAACCYT (SEQ ID NO:48);
GCTGAGGGTGTGTMC (SEQ ID NO:49); ACCAGTGCGGATGKA (SEQ ID NO:50);
CTGGCTGTGTGAAYG (SEQ ID NO:51); GTGACTTGGCCACRT (SEQ ID NO:52);
ACTTCCTAACAAGRC (SEQ ID NO:53); GCAGAGAAAGAGGYC (SEQ ID NO:54);
GCTACTTCCGGGCRG (SEQ ID NO:55); GCTGCAGGGCCCCYG (SEQ ID NO:56);
GGGTGCAGGTGGGKC (SEQ ID NO:57); CTGCCACTGGGGGMC (SEQ ID NO:58);
CTTTGTGTGGCAAMG (SEQ ID NO:59); TAGCCTTCATGCCKT (SEQ ID NO:60);
GGGCTCCAAAGGARA (SEQ ID NO:61); AAGTCCCTGCCATYT (SEQ ID NO:62);
CAGGGACTTGACCYG (SEQ ID NO:63); GAGCAGCAGGGGCRG (SEQ ID NO:64);
CTGCGCAGGTAGGSA (SEQ ID NO:65); GGGAAGTGCGCCTSC (SEQ ID NO:66);
CACAGGGATGTTGMG (SEQ ID NO:67); CTGTACTTTAAGCKC (SEQ ID NO:68);
CTCACCATCCCCCSC (SEQ ID NO:69); GTTATATGAGGAGSG (SEQ ID NO:70);
TAACTGCTCACCTKC (SEQ ID NO:71); CTTTAGCTGCCCGMA (SEQ ID NO:72);
ACTGCCCTGTTTCYG (SEQ ID NO:73); GAGGGGCGAGGGCRG (SEQ ID NO:74);
CACTCTGTCTCTTRA (SEQ ID NO:75); AGAGTTTAGAGATYA (SEQ ID NO:76);
TCAGGGCGTCCACYG (SEQ ID NO:77); and ACGGTGTGGGGGCRG (SEQ ID NO:78).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting PLTP gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AL008726.1

CTTTGTGCCT (SEQ ID NO:79); CAACCAGGTC (SEQ ID NO:80);
TCCCTATCCC (SEQ ID NO:81); GGACTGGGAA (SEQ ID NO:82);
ACCTGAGGGA (SEQ ID NO:83); GGTCTCCCGG (SEQ ID NO:84);
AGTCGCATCG (SEQ ID NO:85); GCTCAGAGAG (SEQ ID NO:86);
CCAGGCTAGT (SEQ ID NO:87); CATAAGTGGG (SEQ ID NO:88);
GCAGGCGCAC (SEQ ID NO:89); GAACTCTGCA (SEQ ID NO:90);
GCTGAGTGGG (SEQ ID NO:91); CCACGCCCCA (SEQ ID NO:92);
CCTGAGTTTC (SEQ ID NO:93); ATGAAGAAAC (SEQ ID NO:94);
GCATTTCATA (SEQ ID NO:95); CTATCAACTC (SEQ ID NO:96);
GTTGATAGGA (SEQ ID NO:97); CTACCAAACC (SEQ ID NO:98);
GAGGGTGTGT (SEQ ID NO:99); AGTGCGGATG (SEQ ID NO:100);
GCTGTGTGAA (SEQ ID NO:101); ACTTGGCCAC (SEQ ID NO:102);
TCCTAACAAG (SEQ ID NO:103); GAGAAAGAGG (SEQ ID NO:104);
ACTTCCGGGC (SEQ ID NO:105); GCAGGGCCCC (SEQ ID NO:106);
TGCAGGTGGG (SEQ ID NO:107); CCACTGGGGG (SEQ ID NO:108);
TGTGTGGCAA (SEQ ID NO:109); CCTTCATGCC (SEQ ID NO:110);
CTCCAAAGGA (SEQ ID NO:111); TCCCTGCCAT (SEQ ID NO:112);
GGACTTGACC (SEQ ID NO:113); CAGCAGGGGC (SEQ ID NO:114);
CGCAGGTAGG (SEQ ID NO:115); AAGTGCGCCT (SEQ ID NO:116);

AGGGATGTTG (SEQ ID NO:117); TACTTTAAGC (SEQ ID NO:118);
ACCATCCCCC (SEQ ID NO:119); ATATGAGGAG (SEQ ID NO:120);
CTGCTCACCT (SEQ ID NO:121); TAGCTGCCCG (SEQ ID NO:122);
GCCCTGTTTC (SEQ ID NO:123); GGGCGAGGGC (SEQ ID NO:124);
TCTGTCTCTT (SEQ ID NO:125); GTTTAGAGAT (SEQ ID NO:126);
GGGCGTCCAC (SEQ ID NO:127); and GTGTGGGGGC (SEQ ID NO:128).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

PLTP genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized PLTP genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the PLTP gene in an individual. As used herein, the terms "PLTP genotype" and "PLTP haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the PLTP gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the PLTP gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25 in the two copies to assign a PLTP genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair

at each of PS1-PS25.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the PLTP gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If a PLTP gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the PLTP gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25 in that copy to assign a PLTP haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the PLTP gene or fragment such as one of the methods described above for preparing PLTP isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two PLTP gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional PLTP clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the PLTP gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS25 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the PLTP haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's PLTP gene, the phased sequence of nucleotides present at each of PS1-PS25. The present invention also contemplates that typically only a subset of PS1-PS25 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a PLTP haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting

of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25 in each copy of the PLTP gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS25 in each copy of the PLTP gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the PLTP gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target

sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the PLTP gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989;

Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's PLTP haplotype pair is predicted from its PLTP genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a PLTP genotype for the individual at two or more PLTP polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing PLTP haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the PLTP haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be

due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a PLTP haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a PLTP genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel PLTP polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for PLTP genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a PLTP genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each

individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular PLTP genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that PLTP genotype, haplotype, or haplotype pair. Preferably, the PLTP genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting PLTP or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a PLTP genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one

or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the PLTP gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and PLTP genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their PLTP genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the PLTP gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between PLTP haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the PLTP gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the

skilled artisan that predicts clinical response as a function of PLTP genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the PLTP gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the PLTP gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying PLTP genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the PLTP gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant PLTP gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25. Similarly, the nucleotide sequence of a variant fragment of the PLTP gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the PLTP gene, which is defined by haplotype 24, (or other reported PLTP sequences) or to portions of the reference sequence (or other reported PLTP sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of adenine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, cytosine at PS5, guanine at PS6, guanine at PS7, adenine at PS8, thymine at PS9, guanine at PS10, adenine at PS11, thymine at PS12, guanine at PS13, adenine at PS14, thymine at PS15, cytosine at PS16, guanine at PS17, thymine at PS18, guanine at PS19, adenine at PS20, guanine at PS21, thymine at PS22, thymine at PS23, adenine at PS24 and cytosine at PS25. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the PLTP gene which is defined by any one of haplotypes 1- 23 and 25 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the PLTP gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

PLTP isogenes may be isolated using any method that allows separation of the two "copies" of the PLTP gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides PLTP genome anthologies, which are collections of PLTP isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A PLTP genome anthology may comprise individual PLTP isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the PLTP isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred PLTP genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded PLTP protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant PLTP sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell

using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the PLTP gene will produce PLTP mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a PLTP cDNA comprising a nucleotide sequence which is a polymorphic variant of the PLTP reference coding sequence shown in Figure 2. Thus, the invention also provides PLTP mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO: 2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 47, adenine at a position corresponding to nucleotide 371, adenine at a position corresponding to nucleotide 849 and cytosine at a position corresponding to nucleotide 1452. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a PLTP isogene defined by haplotypes 1- 23 and 25. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized PLTP cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a PLTP gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the PLTP polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the PLTP gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are

complementary to the sense strand of the PLTP genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular PLTP protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the PLTP isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular PLTP isogene. Expression of a PLTP isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of PLTP mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of PLTP mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference PLTP amino acid sequence shown in Figure 3. The location of a variant amino acid in a PLTP polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig.3). A PLTP protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 16 in isoform 1 and tyrosine at a position corresponding to amino acid position 124 in isoform 1. The invention specifically excludes amino acid sequences identical to those previously identified for PLTP, including SEQ ID NO:3, and previously described fragments thereof. PLTP protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a PLTP protein variant of the invention is encoded by an

isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variants of PLTP(Isoform 1)

Polymorphic Variant Number	Amino Acid Position and Identities	
	16	124
1	H	Y
2	R	S
3	R	Y

The invention also includes PLTP peptide variants, which are any fragments of a PLTP protein variant that contain one or more of the amino acid variations shown in Table 2. A PLTP peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such PLTP peptide variants may be useful as antigens to generate antibodies specific for one of the above PLTP isoforms. In addition, the PLTP peptide variants may be useful in drug screening assays.

A PLTP variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant PLTP genomic and cDNA sequences as described above. Alternatively, the PLTP protein variant may be isolated from a biological sample of an individual having a PLTP isogene which encodes the variant protein. Where the sample contains two different PLTP isoforms (i.e., the individual has different PLTP isogenes), a particular PLTP isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular PLTP isoform but does not bind to the other PLTP isoform.

The expressed or isolated PLTP protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the PLTP protein as discussed further below. PLTP variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant PLTP gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric PLTP protein. The non-PLTP portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the PLTP and non-PLTP portions so that the PLTP protein may be cleaved and purified away from the non-PLTP portion.

An additional embodiment of the invention relates to using a novel PLTP protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known PLTP protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The

PLTP protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a PLTP variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the PLTP protein(s) of interest and then washed. Bound PLTP protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel PLTP protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the PLTP protein.

In yet another embodiment, when a particular PLTP haplotype or group of PLTP haplotypes encodes a PLTP protein variant with an amino acid sequence distinct from that of PLTP protein isoforms encoded by other PLTP haplotypes, then detection of that particular PLTP haplotype or group of PLTP haplotypes may be accomplished by detecting expression of the encoded PLTP protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel PLTP variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The PLTP protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the PLTP protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the PLTP isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the PLTP protein variant from solution as well as react with PLTP protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect PLTP protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel PLTP protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the PLTP protein variant and the antibody is detected. As described, suitable

immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of PLTP may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the PLTP gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into PLTP protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired PLTP isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the PLTP isogene is introduced into a cell in such a way that it recombines with the endogenous PLTP gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired PLTP gene polymorphism. Vectors for the introduction of genes both

for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the PLTP isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the PLTP isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant PLTP gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the PLTP isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human PLTP isogene and producing human PLTP protein can be used as biological models for studying diseases related to abnormal PLTP expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel PLTP isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel PLTP isogenes; an antisense oligonucleotide directed against one of the novel PLTP isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel PLTP isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel PLTP isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule

antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the PLTP gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The PLTP polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include

detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the PLTP gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the PLTP gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in the indicated GenBank Accession Number.

PCR Primer Pairs

GenBank Acc No.	Fragment No.	Forward Primer	Reverse Primer (complement of)	PCR Product
AL008726.1	Fragment 1	1021-1043	1642-1623	622 nt
	Fragment 2	1346-1367	1840-1821	495 nt
	Fragment 3	1893-1913	2316-2297	424 nt
	Fragment 4	2055-2075	2523-2500	469 nt
	Fragment 5	2135-2157	2718-2698	584 nt
	Fragment 6	3409-3432	4035-4013	627 nt
	Fragment 7	3760-3779	4348-4326	589 nt
	Fragment 8	5564-5587	6028-6008	465 nt
	Fragment 9	5712-5734	6157-6135	446 nt
	Fragment 10	7166-7187	7658-7636	493 nt
	Fragment 11	8323-8345	8892-8872	570 nt
	Fragment 12	8579-8599	9066-9046	488 nt
	Fragment 13	10809-10832	11449-11429	641 nt
	Fragment 14	11176-11196	11685-11662	510 nt
	Fragment 15	13611-13633	14027-14006	417 nt
	Fragment 16	13816-13838	14286-14263	471 nt
	Fragment 17	14360-14382	14988-14966	629 nt
	Fragment 18	14613-14633	15117-15096	505 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec.
72°C - 45 sec. } 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in the indicated GenBank Accession Number. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

	Sequencing Primer Pairs		
GenBank Acc No.	Fragment No.	Forward Primer	Reverse Primer (complement of)
AL008726.1	Fragment 1	1066-1086	1614-1596
	Fragment 2	1385-1404	1763-1744
	Fragment 3	1915-1933	2295-2276
	Fragment 4	2095-2114	2496-2477
	Fragment 5	2286-2305	2684-2666
	Fragment 6	3471-3490	3958-3939
	Fragment 7	3815-3834	4319-4300
	Fragment 8	5621-5640	5988-5969
	Fragment 9	5754-5773	6120-6101
	Fragment 10	7205-7225	7587-7568
	Fragment 11	8431-8452	8870-8851
	Fragment 12	8633-8652	9010-8991
	Fragment 13	10941-10962	11379-11361
	Fragment 14	11213-11232	11603-11583
	Fragment 15	13648-13667	13975-13956
	Fragment 16	13877-13896	14214-14195
	Fragment 17	14446-14465	14874-14853
	Fragment 18	14680-14699	15065-15046

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the PLTP gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the PLTP Gene

Polymorphic Site Number	PolyId	Nucleotide Position	Reference Allele	Variant Allele
PS1	12749	1128(Acc#AL008726.1)	G	A
PS2	12852	1632(Acc#AL008726.1)	G	C
PS3	12851	1760(Acc#AL008726.1)	G	C
PS4	12847	2047(Acc#AL008726.1)	C	T
PS5	12848	2146(Acc#AL008726.1)	T	C
PS6	12753	2279(Acc#AL008726.1)	A	G
PS7	12755	2634(Acc#AL008726.1)	A	G
PS8	12761	3840(Acc#AL008726.1)	C	A
PS9	12764	3884(Acc#AL008726.1)	C	T
PS10	12765	3897(Acc#AL008726.1)	A	G
PS11	12766	4055(Acc#AL008726.1)	C	A
PS12	12768	4308(Acc#AL008726.1)	C	T
PS13	12786	7462(Acc#AL008726.1)	A	G
PS14	12803	8710(Acc#AL008726.1)	G	A
PS15	12805	8955(Acc#AL008726.1)	G	T
PS16	12813	10907(Acc#AL008726.1)	A	C
PS17	12811	11040(Acc#AL008726.1)	A	G
PS18	12810	11058(Acc#AL008726.1)	C	T
PS19	12806	11424(Acc#AL008726.1)	C	G
PS20	12814	11486(Acc#AL008726.1)	C	A
PS21	12834	13913(Acc#AL008726.1)	C	G
PS22	12833	13936(Acc#AL008726.1)	G	T
PS23	12837	13982(Acc#AL008726.1)	C	T
PS24	12842	14559(Acc#AL008726.1)	G	A
PS25	12844	14714(Acc#AL008726.1)	T	C

EXAMPLE 2

This example illustrates analysis of the PLTP polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4(Part1). Genotypes and Haplotype Pairs Observed for PLTP Gene

Genotype	Polymorphic Sites											
Number	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP	Pair
1	G	G	G	C	T	A	A	C	C	A	24	24
2	G	G	G	C	C	A	A	A	C	G	10	10
3	G	C	C	C	T	A	A	C	C	A	4	4
4	G	G	G	C	C	A	A	A	C	G	14	14
5	G	G	G	C	C	A	A	A	C	G	10	13
6	G	G	G	C	T	A	A	A/C	C	G/A	18	25
7	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	14	4
8	G	G	G	C	T/C	A	A	C/A	C	A/G	28	13
9	G	C/G	C	C	T/C	A	A	C/A	C	A/G	5	9
10	G	G/C	G/C	C	T	A	A	A/C	C	G/A	18	7
11	G	G/C	G/C	C	T	A	A	A/C	C	G/A	17	8
12	G	G	G	C	T	A	A	A/C	C	G/A	18	28
13	G	C/G	C/G	C	T	A	A/G	C	C	A	4	31
14	G	G	G	C	T/C	A/G	A	C/A	C	A	23	16
15	G	G	G	C	T	A	A	C/A	C	A/G	26	19
16	G	G	G	C	C/T	A	A	A/C	C	G/A	10	26
17	G	G	G	C	C	A	A	A	C	G	14	11
18	G	G	G	C	T/C	A	A	C	C	A	24	15
19	G	G	G	C	T	A	A	A/C	C	G/A	18	26
20	G	C/G	C/G	C	T	A	A	C	C	A	4	29
21	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	10	6
22	G	G	G	C	T	A	A/G	C	C	A	24	32
23	G	G	G	C	C/T	A	A	A	C	G	10	20
24	G	G	G	C	T	A	A	A	C	G	17	21
25	G	G	G	C	C/T	A	A	A/C	C	G/A	10	27
26	G	G	G	C	T/C	A	A	A	C	G	18	14
27	G	G	G	C	C	A	A	A	C	G	10	12
28	G	G	G	C	C/T	A	A	A/C	C	G/A	10	28
29	G	G	G	C	C/T	A	A	A/C	C	G/A	10	22
30	G/A	G	G	C	C	A	A	A	C	G	10	1
31	G	G	G	C	C	A	A	A	C	G	10	14
32	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	10	5
33	G	C	C	C	T/C	A	A	C/A	C	A/G	4	2
34	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	10	8
35	G	G	G	C	C	A	A	A	C	G	10	11
36	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	10	4
37	G	G	G	C	C/T	A	A	A/C	C	G/A	14	30
38	G	G	G	C	C/T	A	A	A	C	G	10	17
39	G	G	G	C	C/T	A	A	A	C	G	14	17
40	G	G	G	C	C/T	A	A	A/C	C	G/A	10	30
41	G	G	G	C	T	A	A	A/C	C	G/A	17	30
42	G	G	G	C	C/T	A	A	A/C	C	G/A	10	24
43	G	G	G	C	T	A	A	A/C	C	G/A	18	24
44	G	G/C	G/C	C	C/T	A	A	A/C	C	G/A	10	3
45	G	G	G	C	C/T	A	A	A	C	G	10	18

Table 4(Part2). Genotypes and Haplotype Pairs Observed for PLTP Gene

Genotype Number	Polymorphic Sites PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20	HAP Pair	
1	C	C	A	G	G	A	A	C	C	C	24	24
2	C	C	A	G	G	A	A	C	C	A	10	10
3	C	C	A	G	G	A	A	C	C	A	4	4
4	C	C	A	G	G	C	A	C	C	C	14	14
5	C	C	A	G	G	A/C	A	C	C	A	10	13
6	C	C	A	G	G	A	A	C	C	C	18	25
7	C	C	A	G	G	C/A	A	C	C	C/A	14	4
8	C	C	A	G	G	C	A	C	C	C/A	28	13
9	C	C	A	G	G	A	A	C	C	C	5	9
10	C	C	A	G	G/T	A	A	C	C	C/A	18	7
11	C	C	A	G	G/T	A	A	C	C	A/C	17	8
12	C	C	A	G	G	A/C	A	C	C	C	18	28
13	C/A	C	A	G	G	A/C	A	C/T	C	A	4	31
14	C	C	A	G	G	A/C	A	C	C	A	23	16
15	C	C	A	G	G	C	A	C	C	A	26	19
16	C	C	A	G	G	A/C	A	C	C	A	10	26
17	C	C	A	G	G	C/A	A	C	C	C	14	11
18	C	C	A	G	G	A	A	C	C	C/A	24	15
19	C	C	A	G	G	A/C	A	C	C	C/A	18	26
20	C	C	A	G	G	A/C	A/G	C	C/G	A	4	29
21	C	C	A	G	G	A/C	A	C	C	A	10	6
22	C	C	A	G	G	A/C	A	C/T	C	C/A	24	32
23	C	C/T	A	G	G	A/C	A	C	C	A	10	20
24	C	C/T	A/G	G	G	A/C	A	C	C	A	17	21
25	C	C	A	G	G	A/C	A	C	C	A/C	10	27
26	C	C	A	G	G	A/C	A	C	C	C	18	14
27	C	C	A	G	G	A/C	A	C	C	A	10	12
28	C	C	A	G	G	A/C	A	C	C	A/C	10	28
29	C	C	A	G/A	G	A/C	A	C	C	A	10	22
30	C	C	A	G	G	A/C	A	C	C	A/C	10	1
31	C	C	A	G	G	A/C	A	C	C	A/C	10	14
32	C	C	A	G	G	A	A	C	C	A/C	10	5
33	C	C	A	G	G	A	A	C	C	A	4	2
34	C	C	A	G	G/T	A	A	C	C	A/C	10	8
35	C	C	A	G	G	A	A	C	C	A/C	10	11
36	C	C	A	G	G	A	A	C	C	A	10	4
37	C	C/T	A	G	G	C	A	C/T	C	C/A	14	30
38	C	C	A	G	G	A	A	C	C	A	10	17
39	C	C	A	G	G	C/A	A	C	C	C/A	14	17
40	C	C/T	A	G	G	A/C	A	C/T	C	A	10	30
41	C	C/T	A	G	G	A/C	A	C/T	C	A	17	30
42	C	C	A	G	G	A	A	C	C	A/C	10	24
43	C	C	A	G	G	A	A	C	C	C	18	24
44	C	C	A	G	G	A	A	C	C	A	10	3
45	C	C	A	G	G	A	A	C	C	A/C	10	18

Table 4(Part3). Genotypes and Haplotype Pairs Observed for PLTP Gene

Genotype Number	Polymorphic Sites PS21	PS22	PS23	PS24	PS25	HAP Pair	
1	C	G	C	G	T	24	24
2	C	G	C	G	T	10	10
3	C	G	C	G	T	4	4
4	C	G	C	G	T	14	14
5	C	G	C/T	G	T	10	13
6	C	G/T	C	G	T	18	25
7	C	G	C	G	T	14	4
8	C	G	C/T	G	T	28	13
9	C	G	C	G	T	5	9
10	C	G/T	C	G	T	18	7
11	C	G	C	G	T	17	8
12	C	G	C	G	T	18	28
13	C	G	C	G	T	4	31
14	C	G	C	G/A	T	23	16
15	C	G	C	G	T	26	19
16	C	G	C	G	T	10	26
17	C	G	C	G	T	14	11
18	C	G	C	G	T	24	15
19	C	G	C	G	T	18	26
20	C	G	C	G	T	4	29
21	C	G	C	G	T	10	6
22	C	G	C	G	T	24	32
23	C/G	G	C	G	T	10	20
24	C/G	G	C	G	T	17	21
25	C	G	C	G/A	T	10	27
26	C	G	C	G	T	18	14
27	C	G	C	G	T	10	12
28	C	G	C	G	T	10	28
29	C	G	C	G	T	10	22
30	C	G	C	G	T	10	1
31	C	G	C	G	T	10	14
32	C	G	C	G	T	10	5
33	C	G	C	G	T	4	2
34	C	G	C	G	T	10	8
35	C	G	C	G	T	10	11
36	C	G	C	G	T	10	4
37	C	G	C	G	T	14	30
38	C	G	C	G	T	10	17
39	C	G	C	G	T	14	17
40	C	G	C	G	T	10	30
41	C	G	C	G	T	17	30
42	C	G	C	G	T	10	24
43	C	G	C	G	T	18	24
44	C	G	C	G	T/C	10	3
45	C	G	C	G	T	10	18

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one

of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 32 human PLTP haplotypes shown in Table 5 below.

Table 5(Part1). Haplotypes Identified in the PLTP Gene

Haplotype Number	Polymorphic Sites PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16	PS 17	PS 18	PS 19	PS 20
1	A	G	G	C	C	A	A	A	C	G	C	C	A	G	G	C	A	C	C	C
2	G	C	C	C	C	A	A	A	C	G	C	C	A	G	G	A	A	C	C	A
3	G	C	C	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	A
4	G	C	C	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	A
5	G	C	C	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	C
6	G	C	C	C	T	A	A	C	C	A	C	C	A	G	G	C	A	C	C	A
7	G	C	C	C	T	A	A	C	C	A	C	C	A	G	T	A	A	C	C	A
8	G	C	C	C	T	A	A	C	C	A	C	C	A	G	T	A	A	C	C	C
9	G	G	C	C	C	A	A	A	C	G	C	C	A	G	G	A	A	C	C	C
10	G	G	G	C	C	A	A	A	C	G	C	C	A	G	G	A	A	C	C	A
11	G	G	G	C	C	A	A	A	C	G	C	C	A	G	G	A	A	C	C	C
12	G	G	G	C	C	A	A	A	C	G	C	C	A	G	G	C	A	C	C	A
13	G	G	G	C	C	A	A	A	C	G	C	C	A	G	G	C	A	C	C	A
14	G	G	G	C	C	A	A	A	C	G	C	C	A	G	G	C	A	C	C	C
15	G	G	G	C	C	A	A	C	C	A	C	C	A	G	G	A	A	C	C	A
16	G	G	G	C	C	G	A	A	C	A	C	C	A	G	G	C	A	C	C	A
17	G	G	G	C	T	A	A	A	C	G	C	C	A	G	G	A	A	C	C	A
18	G	G	G	C	T	A	A	A	C	G	C	C	A	G	G	A	A	C	C	C
19	G	G	G	C	T	A	A	A	C	G	C	C	A	G	G	C	A	C	C	A
20	G	G	G	C	T	A	A	A	C	G	C	T	A	G	G	C	A	C	C	A
21	G	G	G	C	T	A	A	A	C	G	C	T	G	G	G	C	A	C	C	A
22	G	G	G	C	T	A	A	C	C	A	C	C	A	A	G	C	A	C	C	A
23	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	A
24	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	C
25	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	A	A	C	C	C
26	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	C	A	C	C	A
27	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	C	A	C	C	C
28	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	C	A	C	C	C
29	G	G	G	C	T	A	A	C	C	A	C	C	A	G	G	C	G	C	G	A
30	G	G	G	C	T	A	A	C	C	A	C	T	A	G	G	C	A	T	C	A
31	G	G	G	C	T	A	G	C	C	A	A	C	A	G	G	C	A	T	C	A
32	G	G	G	C	T	A	G	C	C	A	C	C	A	G	G	C	A	T	C	A

Table 5(Part2). Haplotypes Identified in the PLTP Gene

Haplotype Number	Polymorphic Sites PS 21	PS 22	PS 23	PS 24	PS 25
1	C	G	C	G	T
2	C	G	C	G	T
3	C	G	C	G	C
4	C	G	C	G	T
5	C	G	C	G	T
6	C	G	C	G	T
7	C	T	C	G	T
8	C	G	C	G	T
9	C	G	C	G	T
10	C	G	C	G	T
11	C	G	C	G	T
12	C	G	C	G	T
13	C	G	T	G	T
14	C	G	C	G	T
15	C	G	C	G	T
16	C	G	C	A	T
17	C	G	C	G	T
18	C	G	C	G	T
19	C	G	C	G	T
20	G	G	C	G	T
21	G	G	C	G	T
22	C	G	C	G	T
23	C	G	C	G	T
24	C	G	C	G	T
25	C	T	C	G	T
26	C	G	C	G	T
27	C	G	C	A	T
28	C	G	C	G	T
29	C	G	C	G	T
30	C	G	C	G	T
31	C	G	C	G	T
32	C	G	C	G	T

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the phospholipid transfer protein (PLTP) gene of an individual which comprises determining whether the individual has one of the PLTP haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS25 on at least one copy of the individual's PLTP gene.

3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS25 on both copies of the individual's PLTP gene.

4. A method for genotyping the phospholipid transfer protein (PLTP) gene of an individual, comprising determining for the two copies of the PLTP gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25.

5. The method of claim 4, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the PLTP gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

6. The method of claim 4, which comprises determining for the two copies of the PLTP gene present in the individual the identity of the nucleotide pair at each of PS1-PS25.

7. A method for haplotyping the phospholipid transfer protein (PLTP) gene of an individual which

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

9. A method for predicting a haplotype pair for the phospholipid transfer protein (PLTP) gene of an individual comprising:

(a) identifying a PLTP genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS1-PS25;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the data in Table 4; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS25.

11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the phospholipid transfer protein (PLTP) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-32 shown in Table 5 and the haplotype pair is selected from the haplotype pairs shown in Table 4, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

12. The method of claim 11, wherein the trait is a clinical response to a drug targeting PLTP.

13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the phospholipid transfer protein (PLTP) gene at a polymorphic site selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25.

14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the PLTP gene at a region containing the polymorphic site.

15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-28, the complements of SEQ ID NOS:4 -28, and SEQ ID NOS:29-78.

16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

17. The composition of claim 16, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:79- 128.

18. A kit for genotyping the PLTP gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24 and PS25.

19. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the phospholipid transfer protein (PLTP) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises a PLTP isogene defined by a haplotype selected from the group consisting of haplotypes 1-32 in Table 5; and

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

20. The isolated polynucleotide of claim 19, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

21. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 19, wherein the organism expresses a PLTP protein encoded by the first nucleotide sequence.

22. The recombinant organism of claim 21, which is a nonhuman transgenic animal.

23. The isolated polynucleotide of claim 19, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the PLTP gene, the fragment comprising one or more polymorphisms selected from the group consisting of adenine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, cytosine at PS5, guanine at PS6, guanine at PS7, adenine at PS8, thymine at PS9, guanine at PS10, adenine at PS11, thymine at PS12, guanine at PS13, adenine at PS14, thymine at PS15, cytosine at PS16, guanine at PS17, thymine at PS18, guanine at PS19, adenine at PS20, guanine at PS21, thymine at PS22, thymine at PS23, adenine at PS24 and cytosine at PS25.

24. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the PLTP cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of a PLTP isogene defined by one of the haplotypes shown in Table 5.

25. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 24, wherein the organism expresses a phospholipid transfer protein (PLTP) protein encoded by the polymorphic variant sequence.

26. The recombinant organism of claim 25, which is a nonhuman transgenic animal.

27. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the PLTP protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 5.

28. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 27.

29. A method for screening for drugs targeting the isolated polypeptide of claim 27 which comprises contacting the PLTP polymorphic variant with a candidate agent and assaying for binding activity.

30. A computer system for storing and analyzing polymorphism data for the phospholipid transfer protein gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.

31. A genome anthology for the phospholipid transfer protein (PLTP) gene which comprises PLTP isogenes defined by any one of haplotypes 1-32 shown in Table 5.

POLYMORPHISMS IN THE PLTP GENE (Accession No. AL008726.1)

```
GATCACTTGA GGGCAGGAGT TCAAGACCAG CCTGGCCAAA ATGGCAAAAC
CCTGTCTCCA CTAAAAATAC AAAAATTAGC CAGGTGTGGT GGCACATGGC      100
TGTAGTCCCA GCTACTTGGG AGGCTGAGGC AGGAGAATCA CTTGAACCCA
GGAGGCAGAG GTTGCAGTGA GCCAAGATCA CACCACCGCA CTGCAGCCTG      200
GTGACAGAGC ACGACTGTGT CTCAAAAAAA TTAATTAATT AATTAAATAA
AAAAGGAAAT GGAACTATTT TTGTGAATCT GTGGATTATA TCAGAAAAAA      300
AAGACACAAT GGGGAAAGTC CTAGGCAAAT CAGGATGAGT TAGTCATCCT
TCCTAGATGA GTGTTTGGTG CTAAATACAT GCTCAGCAGA CATGATTATT      400
GCTTCCCCTT TCTTTCGTCC ATTTGGCAAC AAAAAGGTGG CAAGCACCCA
CTCTGTGCCC TGTCCTAGGG TCCGGGAACC CTGTAAGCAG TAGATGGAGG      500
TGGGGGTGGG GGTGGGGGCG GGGATGCTGT TCAGAGCACC TTGCTCCAAG
GGTTCATTAA AAAATCCACC AGTGGACCGG GCGCGGTGGC TCATGCCTTT      600
AATCCCAGCA CTTTGGGAGG CCGAGGCGGG CGGATCACAA GGTCAGGAGA
TCGAGACCAT CCTGGCTAAC ACGGTGAAAC CCCGTCTCTA CTAAAAATAC      700
AAAAAAAATT AGCCGGGCGT GGCAGCGTGC GCCTGTAGCC CCAGCTGCTG
GGGAGGCTGA GGCAGGAGAA TGGCGTGAAC CCGGGAGGCG GGGCTTGCAG      800
GGAGCCGAGA TCGCGCCACT GCACTCCAGC CTGGGGGACA GAGCGAGACT
CCGTCTCAAA AAATAAAAAT AAAAAAAATA GAAAAAACAA TCCACCAGCC      900
ACGATAAATG GCAGACCTCC TTCTGATTTC AGCCGGTGTG GTATGTTCCT
GGGCTGACAG CACTTGTCTA GTCTTGCTTT CCCAAGTGGG AAAGGTCTCT     1000
GGGACCTTAA GGTCCCCAGG TGGTGACACA GAGACAGGTA GGGGGGCCCA
TAGCAAAGCC AGGCAAGGAG GTCCCGAGAT GATTGTGGGT GGCAGGGAAA     1100
GAAAAAATAT TCCTTGACTT TGTGCCTGGA CCTGGTTGTA ATAAAGGCCC
                          A
AAGAGGTAGT TCCTATCATC GTGCACATTT CGCTGAAGGA AGAAACTGAG     1200
GGTCAGTGAC CCAAGTGAAG TGACTTGCCC AAGATCATGC AGGAAGACAT
GGATAATTGT AATTTGAACC AAGGTCCCAG CAAAGTGGGA TTGTTGGGGC     1300
TGAGTGGGCC GGCTCCTGCA TTTCCTTCCC TCTCCCTGGG CTTGGGTCTC
CCACTTGTCC AGACAGCGGC CGGGCTTGTC ACGGGGCTCT GTGCAGCCTT     1400
TTCCACTCTC CCGGCTGCCA GCGTCCCGCC CCGTCCCCTC CCAGCCCCCA
AGGGAGGAGG GGAGAGCTGC AGAGAGGAGG AGGGGTCGGG GAGGCCGGCT     1500
TTATAAAGGC GGCTGGAACA ACCCTGCCCG CCAGACCCCG TCGCCCGGAT
     [EXON 1: 1538..
CCCCTGAGCT GCCCGCCATC CCACGTGACC GCGCCGCCCC CCAGCTCCAC     1600
CGCTGAGTGA GTTGGGGCGC GTCCCTATCC CGTTCCCAGT CCCTCGCCGA
                                   C
                ..1606]
GCCGCGTGGT GCACTGAAGC CGGCGTGGGG AGGAGAGCCG ATGAAGGAGA     1700
ACGCTAACAT GGGGGCTCCA GGCAGAATCT CTAATGGGAG AGATTTAGGA
CCTGAGGGAG CCGGGAGACC CGGGAGCCCA CGGTCTGGTC GGCCACCTCC     1800
        C
TCTCCTCCCC GGGCGCGAGG CCTATGAGAC CGCAAGCCAC CCTCCAAGAA
TGCGCGTGCA GTCGTTGCCG TGGCAACGTG CAAGCCTGCA TGGGCCCACC     1900
GCTAAAGGAA AGGGGAGGTT GTTGGGGGTG TGGGGGCTGG TCCCCTGGAG
GTCCAAACTC ATCACGCGGC GGTAGCACGA GGGAACTGGG AACGTCCCCT     2000
CCCCCGTGTC CCAAATCCTC TGGGAGCCAG AACGCAAGTC GCATCGCCTC
                                                  T
TCTGAGCCTC AGTTTCCACA TCTGCCATAT GGAGGCAAGA TTATTGGTTC     2100
     [EXON 2: 2053..
CAAGTTCCCC GAAGAGTTAT CTGGAGGGTA ACGATCCAGG CTAGTTCCCA
                                                C
CTTATGGTTG GACGTCCAAG GAATGGGTTT AAGCGGCTCA GATAACTCCT     2200
```

FIGURE 1A

```
GCGACCCCAC CCCACCCCCA GGCCCGCTCG CCATGGCCCT CTTCGGGGCC
CTCTTCCTAG CGCTGCTGGC AGGCGCACAT GCAGAGTTCC CAGGCTGCAA      2300
                              G
GATCCGCGTC ACCTCCAAGG CGCTGGAGCT GGGTAAGGCG CAGGGGCAGA
                ..2332]
CAGGGACGGA CGGGAGCGGA CGGGGTTGGG GTCGCCCCAA CAGTGGGCAC      2400
GGGAGCTAAG TAAGGGTTTC TCTGCTGCTT CAGTGAAGCA GGAGGGGCTG
     [EXON 3: 2434..
CGCTTTCTGG AGCAAGAGCT GGAGACTATC ACCATTCCGG ACCTGCGGGG      2500
CAAAGAAGGC CACTTCTACT ACAACATCTC TGAGTAAGTG GGGGGCGGGG
                ..2533]
CCTCGCGATC CGGGGCGGGG CCTCGGCGGT GGAGGCGGGG CCCGAGAAAT      2600
CTCAGGGTCT CGGGAGACTG GAGGCTGAGT GGGATGGGGC GTGGCCAAAT
                                   G
CGATGAGGCA AGACTTAAGA AACCGGATGG GGCCGAAAAG ATAAAGGCCA      2700
AGCCAGTCAT CATAGGCGGG ACCAGGCAGA TGGGCGTGGC CTTGAAAATG
TAGAGGGGGC CAAGAAGGAA GCAGGACCAG TGGATGTGAG GCGTGGTCGA      2800
AGAACAACAG AAAGGTGGAG TCAAGGAACT TCCCAATAGA TGAGAGTGGA
GTCTCGGGCA CGGGGTGGAG TTAATCGGAG GGGTGGAGAA ACTGAAGGCT      2900
CTAAGTATGG TTTAGGATGA GGGAGGGGGA ATAAGGAATT GAGATGTGAG
ACGGAGCTTA GGTGATGGAG GCAAAGTGAA CGGACTCTAT TAAAAGACCG      3000
GGTGTAGCAA CCCCAGGGTG TGGGAGGGGA GGGGGTGCAA GCAAACAAGA
CTTTTTTTTT TTTTTTTTCT GAGAGGGAGT CTTGTTCTGT TGCCCAGGCT      3100
AGAGTGCAGT GGCGCGATCT TGGCTCACTG CAACCTCCGC ATCCCGAGTT
CAAGCGATTT TCCTGCCTCA GCCTCCCGAG TAGCTAGGAC TACAGGCGCG      3200
TGCCACCACG GTCTGCTAAG TTTTGTATTT TTGGTAGAGA CGGGTTTCAC
TATGTTGGCC AGACTGGTCT CGAACTCCTG ACCTCAAGTG ATCCGCCCTC      3300
CTCGGCCTCC TAAAGTGCTG GGATTACAGG CATGAGCCAC CACCGCACCT
GGCAAACAAA CAAACAAGAC TTTTAAGCGA TGGGGTGAAG CTCATTTTTA      3400
ACACCTGGAT CAGGAAGACC TCCTCTCTCA GCCACCCAAG TTCAACTTTG
AGCCACCCAA GTCTAAGTTG GGCACTTGAG GGACAATGGA AGAATTCATT      3500
CCACCCACTC CCGCTTCCCC AGGCCTGGAC TTGAAAGGGG AGCAGACAAA
TTTCCTGTCG TTGGGGGAAG TTCCCTCTTC TTGGCCCTGG ATCTGACCCT      3600
GAGGCCTCCT GTAGGGTGAA GGTCACAGAG CTGCAACTGA CATCTTCCGA
     [EXON 4: 3615..
GCTCGATTTC CAGCCACAGC AGGAGCTGAT GCTTCAAATC ACCAATGCCT      3700
CCTTGGGGCT GCGCTTCCGG AGACAGCTGC TCTACTGGTT CTTGTAAGGA
                ..3743]
CCCAGCACCC TCAGGGGAGT GGAGAGTGGG TCAGGAGGGC TGGAGGTGCT      3800
GTGGGGCTGT GGGCAGCCCC CTTAGCCTTC CTGAGTTTCC GTTTCTTCAT
                                        A
CTGTGTAACA GAAACAGGCA TATGCATTTC ATACGAGTTG ATAGGAAGGT      3900
                                   T              G
TTGGTAGGAT GTGTGGAAAG AGGTGAAACC TTATTAAGAG GTATGGTGAT
GCATGGGGGC ACTGAAGACT CAATTGAGCT CCTACCTGGA GTGAATATTA      4000
ACCCCCCTGG CAGCTATGAT GGGGGCTACA TCAACGCCTC AGCTGAGGGT
     [EXON 5: 4014..
GTGTCCATCC GCACTGGTCT GGAGCTCTCC CGGGATCCCG CTGGACGGAT      4100
   A
GAAAGTGTCC AATGTCTCCT GCCAGGCCTC TGTCTCCAGA ATGCACGCGG
CCTTCGGGGG AACCTTCAAG TAAGCCCCAG CCCCAACCCC AGCTCACCCT      4200
                ..4169]
TTTCGAGCCC CACAGCTCAT CTGCCTAGCG CAATGCAAAG TGCATAGGCC
TGGGAATCAG ACCAACTGGG TCCAAAGCTC AGCTCTGTCA CTTCCTGGCT      4300
```

FIGURE 1B

```
GTGTGAACGT GGCCAAGTCA CTGAAGCTTG GTGCCTCAAT ATCCTCATTT
      T
TATAAGATCC TATTTTATAA GATCGTTGTC AGGATTATTT GTGTTGGTAT      4400
ATGTAAAAAA TATTTAGAAC AAGGCTGGCT GTGGTGGCTC ATGCCTGAAA
TCCCAGCAAT TTGGGAGGTC TGGGCGGGCG GATCACAAGG TTAACAGATC      4500
GAGATCATCC TGGCCAACAT GGTGAAACCC CATCTCTACT AAAAATACAA
AAATTAGCTG GGCGTGGTGG CACGTGCCTG TAGTCCCAGC TACGTGGGAG      4600
GCTGAGGCAG GAGAATCGCT TGAACCCGGG AGGCGGAGGT TGCAGTGAGC
TGAGATCGTG CCACTGCACT CCAGCCTGGT GACAGAGCAA GATTCCATCT      4700
CAAAAAAAAA AAAAAAAAAA AATTTAGAAC AGCACCTGGC ATATGATAAA
TGTATATTTA AGTGCTGGCC ACTTAGATAA GAGTAACTGT GAGGACTATC      4800
TGATATGCAC ATTTGGCCAT GTACTGGATT CTCAATAATT GGTATCTCTA
AATAATAATA ACAAAAATAA TAATTTTTTT TTTGAGAGGG AGTCTCACTG      4900
TGTCACCCAG CCTGGAGTGC AGTGGCGCCA TCTCGGTTTA CTGCAACCTC
CACCTCCTGG GTTCAAACGA TTCTCCTGCC TCAGCCTCCC AAGTAGCTGG      5000
GACTACAGGC ATGCACCACC ATGCCCAACT ACTTTTTTTG TATTTTTAGT
AGAGACAGGG TTTCGCCATA TTGGCCACGC TGGTCTCAAA CTCCTGATCT      5100
CAGGTGGTCT GTCCGCCTCG GCCTACCAAA GTGCTGGGAT TACAGGCCTG
AGCAACTGTG CCCGGCCTAT TTTTTTATTT TTGCTGCTGA GAAGGGGATA      5200
TTGTTCTTGT TCTGGAAGAG GATGGGGGAG AGAAGATGAT GGATTATAAC
CTGGTGTTGG TACACTGTCA GTATTATTTA TTCATCAAAC AAAGTGTATG      5300
CCAGTCCCTG AGGATGCAGG AAGGAGCTCA GAGCCTATCT GGAGCAAAAA
ACATTTGATA TCAGGGCCTG GAGAGAAGGA CTAAGAAAAT GCTTTGGGGA      5400
TTCAGAGGAG GGGAATCTGG GAAGACTGCC TGGAAGAGGA AGCATCTGAG
CTCAGATGGG AAAATACAGA GTAGGAATGC AGAGGGCGGA AGGGAGGGCA      5500
TCAGTAAGCC GATGGATGTG GGGATGCTCA GAGTGGGTTT GAGGCAATGT
GGGTGGATTC ATTTGACTGA TGGGACCAGA GAGTAGGTCA GGAGGTCTTG      5600
AGAATTAGAC TGGAGAGGGA ATTTGGGCCT GGTTCTTGGA AAACCTACGT
GGAGGAGCTG TTATTCAGGC TGAAGCTGAG GACTCCTACT GCCACTATTT      5700
CCCTAGTCAC TGATTTCTCC TGGACCTGGA CTGGGGTGGA GGCAGGTTCT
GGGCTCATCC GGCCTCTCCT CTCCTCACAG GAAGGTGTAT GATTTTCTCT      5800
     [EXON 6: 5781..
CCACGTTCAT CACCTCAGGG ATGCGCTTCC TCCTCAACCA GCAGGTGTGG
                ..5844]
GCAGCGACAG GTCGCAGGGT GGCAAGGGTG GGCATGCTCT CACTTTGAGA      5900
     [EXON 7: 5880..
AGGCCCTGAC TCTGGCTCCC ACCTCGCAGA TCTGCCCTGT CCTCTACCAC
GCAGGGACGG TCCTGCTCAA CTCCCTCCTG GACACCGTGC CTGGTGAGTT      6000
                ..5993]
GGTGGCGGGT GAGTCTGGGT GTGCAGCTGT CATGCAGCAC CTCAGAGCAG
GCCCCTTCCG AGCCCTGCTG TTGAACAGTC CTGGGTTCAA ATGTGGCCCC      6100
TGGAGCTGAC TTGTTGTGCG ATTTTGGGTG AAGTGCTCAT TTCTGTCTGG
GCAAAGTGCT TGTTTCCTCA CTTGGCTGGC AGTGCCTCCT CCTGGGGTTG      6200
```

```
GGGAAAGGAA CCTCATGCCA TCTTTTTTTT TTTTTTTTTT TTTTGAGATG      6800
GAGTCTCGCT CTGTCACCCA GGCTGGAGTG CAGTGGTGCA ATCCTGGCTC
CCTGCAACCT CCACCTCCCA GGATCAAGCG ATTCTCCTGC CTCAGCCTCC      6900
TGAGTAGCTG GGATTACAGG CGCCCACCAC CATGCCCAAC TAATTTTTGT
ATTTTTAGTA GAGACGGGGT TTCACCGTGG TGGACCAGGC TGGTCTCGAA      7000
CACCTGACCT CAGGTGATCC GCCCGCCTTG ACCTCCCTAA GTGCTGGGAT
TACAGGTGTG AGCCACCATG CGCGGCCCCC CCATGCCATC TTGTTTAGTT      7100
ATTTTTTTGA AACCTCTATA GTGGTAGTAA TAATAATAAC TAACATTAAT
GAGCACCTAA CTACACACCA GGCCCTAAGT GCTTTCCACA TAAAACTAAT      7200
TTGACCCTCA TGACAACCCT GTGAAGGATG TATCCTCATT TTATGGGTGG
GCCTCTGAGC CTGGAGAGGT TAAGCACCTT ATCCCACATC ACACAGCCGT      7300
GAGTGACTTG AGCCCCTTCC TGCAGTGCGC AGTTCTGTGG ACGAGCTTGT
     [EXON 8: 7326..
TGGCATTGAC TATTCCCTCA TGAAGGATCC TGTGGCTTCC ACCAGCAACC      7400
TGGACATGGA CTTCCGGGTG AGCTGCTTGG GCTGGTGTAT GACCTCTGAC
                ..7417]
TTCCTAACAA GACCTCTTTC TCTGCTAAGT TGCAAGATTT TACTTCTTAA      7500
          G
AGCACAGCTC TGATGAGGCC ACTCCCATTA CTGAGAATGC TATGGCTCCC
TGTTACTACA GAATTAGGTT CAGCCTCCTG GGTCTGGCAT TTGAGCCCCT      7600
GTTAATCTTA CCTGGCCCCA CCTTTTATAT CTCATCCTGT TTCTTTGCAG
CTGCTCTTTC CCAAATTGGA CCCTGGCTTC CCACCTCAGG GCCTTTGCTC      7700
ATGCACGGCC TTTCCCCAAC TCCCCCGAAG GAGCTGGACA TAAATTTCTA
TACAAATAAT CTCATTTCAT TCTCTCACAA CCCTGTGAGT CAGGTACTAA      7800
CATCATCCTA ATTTTAGGAA TGTGGAGACT TACTCATAGA GGTGACGTCA
CTTGCCCAAG GCCACATAGC CAGTGTGTAG CCACTACCTT TGGTCTCCCA      7900
GAGGAGTATA GCAGTTTCAG AGTGCAGACT CGGGAGCTAG AATCAGTCGG
GTGCAGTGGC TCATGCCTGT AATCCCAATA CTTTGGGAGG CCCAGGCGAG      8000
AGGATCGCTT GAGCCCAGGA GTTTGAGACC AGCTTGGGCA TCATAGTGAG
ACCTCCTTCT CTACAAAGAG AAAAAGAATG CCTGAAGTTC AAATCCTAGC      8100
TCTGCCACTG ACTAGCTGCA CGGCCCTAAG CATAAGTTTT CAGTTGGCCA
GTTTCCTAAC ATGTAAATGG AAATAATCTA GTAAGAAAAC ACACGTTGTT      8200
ATGCAGATTA AAGAAGTCGA TCTATGTGAA GAGCTTAGAG TACTTGGCAA
AGAAGCATGA GTATGCACAA TATATGTAAA CAGTAGCTGT TGTTACTGCC      8300
TCTAGCTTTC AAGGCCCACT CAGATGTCAC CTCTTCCCTG AAACCTTTCC
TGACCCTCCC TCCTCCAGTC TGACGGGCTC GCCCCCTGGT AGCAGGTGTC      8400
CTCCTTCCAA TGCCTTTTAA GAGATGTGCT TGTGAATCCT CCAAGAACAG
CAGCTGGGTT CCCACCATCC CCCCACAGTG CCTAGCACGG GGCTGGGAGA      8500
CACCTGCTGT CAGTCCAGGT GCCAGGCCAG TCTGCCTTGA CCCTGACTGT
GAATGCCCCA CCCCAGGGGG CCTTCTTCCC CCTGACTGAG AGGAACTGGA      8600
     [EXON 9: 8567..
GCCTCCCCAA CCGGGCAGTG GAGCCCCAGC TGCAGGAGGA AGAGCGGATG
GTGTATGTGG CCTTCTCTGA GTTCTTCTTC GACTCTGCCA TGGAGAGCTA      8700
CTTCCGGGCG GGGGCCCTGC AGCTGTTGCT GGTGGGGGAC AAGGTATGTC
         A
                ..8743]
ATGGCCTGTT TGTGGGATGG GCAAGAGAAG GTCTGTGACA GAGCTCACTC      8800
CCTCACTCCT GATTCCCCTG TTCAGGTGCC CCACGACCTG GACATGCTGC
     [EXON 10: 8826..
TGAGGGCCAC CTACTTTGGG AGCATTGTCC TGCTGGTGAG TGCTGGCGGG      8900
                ..8885]
GGCAGGGATA GGGCCACCTG CACACCAGTG AGACCAAGGA GGGGTGCAGG
TGGGGCCCCC AGTGGCAGCC ACGCAGACAG GGCCCTGGCC TCTTGTCTGT      9000
    T
```

FIGURE 1D

GGCCAATCCC CCTCGCCTTG TCATGGCTGT TCCTGTCTGT GAAGCGGTCT
GCTCCTTCCT GCTCCTTCCT GCCCCATCTC CTTCGCAGAG CTGCTGTGAG 9100
GATTAGTGAG AACCCAAGCT TTGCAAGTGT GAGGGCTTTA ATAATAATTC
ACATATACTT TTGTTTATTT GTGTATTGGG GAGGAGAGTA AGAGTTACAC 9200
ATTTTATTTA TTTTTTTTTT TGAGACAAGA GTCTCACTTT GTCACCCAAG
TTGGAATGCA GTGGCACAAT CTAGGCTCAC TGCAACCTCT GCCTCCCAGG 9300
GCTCAAGCCA TCCTCCTGCT TCAGCCTCTT GAGTAGCTGG AATTACAGGC
TTGTGCCACC ACACCCCGCT AATTTTTATA TTTTGTTGAG ACAGAGTTTC 9400
ATTATGTTGC TCAGGCTGTC CTTGAACTCC TGGGCTCAAG TGATCACCCA
TTTTTGGCCT CCCAAAGTGC TAGGATTACA GGTGTGAGCC ACCGCACCTA 9500
GCAGAATTAC AAATTTTAAA AGAAATAATT CAGTCATTAT CATCATCCAA
TTTATGGCAG GGCAGGGTCT CTATTGTCAG ACTGCCTGGG CTTCAAGCCT 9600
AGTTCTGAAA CTGAGTAACT ACATAACCTC TACAAGTTAC TTAACCTATC
TGTGCCTCAA TTTCCCCATC TCAAAAATGG GGATATGGCT GGGCATGATG 9700
GCTCACGCCT ATACTCTGAG AACTTTGAGA GGTCAAGGTA GGAAAATTGC
TTGAGCCCAG GAGTTTGAGA CCAGCCTGGG CAACACAGTG AGACACTGTC 9800
TCTATTGAAA AATAAATTTT TAGTCAGATC TTTTTTTTTC GGTGGGGGGA
CAAAGTCTCA CTCTGTCACC CAGGCTGGAG TGCAGTGGCG CAATCTTGGC 9900
TCACTGCAAT TGAACCTCCC GGGTTCAAGC AATTCTCTGC CATGACGACT
CAGCCTGAAA AATAAATTTG TAAAAAATAA CAACAACAAA ATTGGCAGTA 10000
ATAATATCGG CCTCACATGG TTATTATGAG GTTTAAAGGA ACCAACACAT
ATACAGTGTT TAGAATCATG TTTGGTGAAG ATTAGGCAGT TGCATCTTAT 10100
AGTATTCCTT CACTGCCTAT AGAAGAAGGC ATAGCAGTTT AGAGAAAGAA
GATATGGCCA GGCACGGTGG CTCACGCCTG TAATCCCAAC ACTGTGGGAG 10200
GCTGAAGCCG GCAGATCACT TGAGCCCAGG AGTTTGAGAC CAGCATGGGC
AACATGGCCA AACCCCGTCT CTACGAAAAC ACAAAAGCAA AAACCAAAAC 10300
AAAAAACAAA TACAAAAATA AGCCAGGCGT GATGGCATGA ACCTGTGGTC
TCAGCTGTTC AGGAGGCTGA TGTGGCAGGA TAATTTGAGC TGAGGAGGTT 10400
GAGGCTGCAG TGAGCCGTGA TCATGCCACT GCACTCCAGC CTGGGCAACA
GAGACCCTGT CTCAAAAAAA AAGGAGAGAG AGAGCAAAAA GATATGGCAA 10500
GGGGTTTGGG TGTGGTGGCT CATGCCTGTA ATCCCAGCAC TTTGGGAGGC
CGAGGTGGGC GAATCACTAG CCTGGCCAAT ATGGCGAAAC CCCGTTTCTA 10600
TTAAAAATAC AAAACTTAGC CGGGCATGGC GGCAGGTGCC TGTAATCTCA
GCTACTTGGG AGGCTGAGGC GGGAGAATCG CCTGAACCTG GGAGACAGAG 10700
TTGCAGTGAG CCGAGATTGT GCCACTGCAC TCCAGCCTGG GCAACAAAGT
GAGACTCCAT ATCAAAAACA GAAGACATGG CAAGGGCTGA GATACTAATG 10800

```
GCAGCTGGAC CTGCGCAGGT AGGCAGGCGC ACTTCCCTGC ACCTCAGTGA
                     G
                ..11418]
TTGGAGGACA ATCACACCTG TCCACAGGGA TGTTGCGCTT AAAGTACAGT      11500
                                    A
GATACCTGTA AGCCACTGAG AGCTGGGGTC TGTCTTTGTT AATGCATGCC
AAAGTGCTTT GGTAAAGTCT GACTCCTCTG CATTCTCACA TGTCCCAGAG      11600
GCAGGCAGGG CAGGTATTAT GAGAGATGAA CAAATGGGTC CAGCAATGTT
AAATAACATG CCCACAGTCA CACAGCAAAG ATCAGAACAT GGGTTTTGAC      11700
AAAAACCTCA GCACAACACA GCTTTGTGAC CTTGAGGCCA TTTTATCTTT
TTTAAAAATG TTATTATTCT TTTTTTTTTT TTTTTGAGAA GGAGTTTCGC      11800
TCTTGTTGCC CAGGCTGGAG TGCAATAGCG TGGTCTTGAC TCACTGCCAC
CTCTGCCTCC CGGGTTTAGG CAATTCTGCC TCAGCTTCCT GAGTAGCTGG      11900
GATTACAGGT GTGCGCCACC ACGCCCGGCT AATTTTTGCA TTTTTAGTAG
AGACAGGGTT TCACCATATT GGTCAGGCTG GTCTCAAACT CCTGACCTCA      12000
GGTGATTCAC CTGCCCCGGC CTCCCAAAGT GCTGGGATTA TAGGTATGAG
CCACTGTGCT TGGCCTTATT TTATTGTATT TATTTATTTT TGAGACAGAG      12100
TCTCACTCTG TTGCCCAAGC TGGAGTGCAG TGGTGCAATC ATAGCTCACT
GCAGCCTTGA CCTCCTGGGC TCAAGTGATC GTCCAGCCTC AGCCTCCCGA      12200
GTAACTGGGA CTGCATGCAT GCACTACCAT GCCTGACTAT TTTATTTTTT
GTAGAGATGG GGTCTCACTT TGTTGTCCAG GCTGGTCTCC AATTCTCGGG      12300
CTCAAGTGAT CGTCCCACCT TAGCCTCCCA AAGTGCTAGG ATTACAGGCG
TGAGCCACCG CACCCAGCCC ATTTTGTCTT TCTAGGCCTC ATTTTCTCCA      12400
TGTGTTTAAT ACAGGCAAAA GTAGCATCTG CCTTCTAGGG CTGTGCTATC
CAATATGGTA GATACTAGCC ACATGTGACT ATTTACATTT AAATTAATTA      12500
TAATTAAATA GTTTTTCAGT TGCACTAAGC CACATTTCAT TAGCCACATG
TGGCTAGTGA CTACCATAGT AGACGGTGCA GATACAAAAC GTTTCACAGA      12600
AAATCTGACT GGACAGTGCT GTTCTAAGGT TAAAGGAGAT AATGTAGGTT
AGGACCCTAG GTTCTAGAGG CAGCTTCTGA CACTTTCCAG CTGTGTGACT      12700
TCAGGCAAAT TACTTAACCT TGCAGGGCCT CATTTTCCTC ATCTATAAAT
TAGTGCAAAT CCTAATAGTG CCCACTCCTT GGGCTTGTTT TGAGGACTAA      12800
ACCAGTTGGT ACTGTATTTT AAAACAGTTT ACTAATATTA ACTTTTTAAA
CACGTTGGCT GGGTGCAGTG GCTCACACCT GTAATCTCAA CACTTTGGGA      12900
GACCAAGGCG GGTAGATCGC TTGAGGCCAG GAGTTCGATA CCAGCCTGAG
CTACAAAGTG AGATCCTGTC TCTACAAAAA AATTTAAAAG TTAGTTGGTC      13000
ATAGTGATGC ACGCCTGTGG TCCCAGCTAC TTGGAAGCCT GAGGTGGGAG
GCTGACTTGA GCCCTGGAGG TCAAAGCCAC AGTGAGTAAA ATTACACCAC      13100
TGCCCTCCAG CCTGGGCAAC ACAGTGAGAC GCTGTCTCAA AAATAAATAA
AAATAAAAAA AGAAAATAAA TATGCTTAGT GTCTGACAGG AACTAAGCAA      13200
TGTGAAGTTT TGTTACATAA AATAAATAAC ACCTGCAAAG CTACCAGCAG
AAGTCTGGCA CATTAGAAGT GCTCAATAAA GGCCGGGCGT GGTGGCTCAC      13300
GCCTGTAATC CCAGCACTTT GGGAGGCCGA GACGGGCGGA TCACAAGGTC
AGGAAATCCG AGACCATCCT GGCTAACACG GTGAAACCCC GTCTCTACTA      13400
AAAACACAAA AAATTAGCTG GGCTTGGTGG CGGGCACCTA TAGTCCCAGT
TACTCGGGAG GCTGAGCCAG GAGAATGGCG TGAACCCGGG AGACGGAGCA      13500
GTGAGCCGAG ATCGCGCCAC TGCACTCCAG CGTGGGCGAC ACAGCGAGAC
TCTGTCTCAA AAAAAAAAAA AAAAAAAAAA AAAAAAGAGG TGCTCAATAA      13600
ATAAAAAGTA GCAGTCTTGG AGAAAAGGAG ACCCAAGTGG CAGCAATGAC
TGCCCCGAGG ATACCTGGAA AGTTGTAGCA CTTGGAAAGG TGCCTAGGTG      13700
GGCATCTGAG CCTGCAACTC AGACACCTCT GGCAGGTAGG CCTGGGTGGC
CTGAGGAAGG GGAGCTTCCA GGCTTCCTGC CCACAGGGCT CCTTTCTTTC      13800
CCGCAGGTTC CGAATCTATT CCAACCATTC TGCACTGGAG TCGCTGGCTG
    [EXON 13: 13807..
              ..13849]
```

FIGURE 1F

```
TGAGTGGGGA AGGGTGGGTG GGCTGGGCTG GGTTCGTGAG ATCCTGACCC      13900
TCACCATCCC CCCCTCCTCA TATAACTGCT CACCTGCGGG CAGCTAAAGG
            G                        T
CAAACCCAAC CCTGACTCAC TGCCCTGTTT CCGCCCTCGC CCCTCCCAGC      14000
                                     T
     [EXON 14: 14000..
TGATCCCATT ACAGGCCCCT CTGAAGACCA TGCTGCAGAT TGGGGTGATG
CCCATGCTCA ATGGTAAGGC TGGGGTGTGA GGATGGAGGA AGAAAGGAGG      14100
                ..14063]
GGTGAACTGG GCGGGCCCAG ACTGAGCGGG GTGCTCCCAC CCACAGAGCG
     [EXON 15: 14147..
GACCTGGCGT GGGGTGCAGA TCCCACTACC TGAGGGCATC AACTTTGTGC      14200
ATGAGGTGGT GACGAACCAT GCGGTGAGTG GGGGCAGGAT GGGGAAGGAG
                ..14223]
GGAGGGCCTT CCCCTCTCTC TATAGCTCCC CAGTGCCCTC CATTACCTGA      14300
CCCCAGCTCC CCCATTACCA GTCGCCCCTC CCATTTTAAT CTTAGACTCC
TAAAAGCCGT TGCACACACT GTGACTTCTT GGAGCACTCT TCTCCCATCC      14400
CAATACCTAC TAATTGTTCA GGTCTTAGCT CAGAGGTCAC CTCCTCCAAG
AAGCCCTCCT TGACACCCAT AGTCTACACT CCTACAGCCT CATTTTAACA      14500
CCTTGACATG GAATTATCCA TTAGATTGAG TTCCTGAAAG ACAAACACTC
TGTCTCTTGA TCTCTAAACT CTGGCCTCAA TAACTGAATG AATAAATGAG      14600
       A
GCTCCCCTCT CCCCAACACA GGGATTCCTC ACCATCGGGG CTGATCTCCA
     [EXON 16: 14622..
CTTTGCCAAA GGGCTGCGAG AGGTGATTGA GAAGAACCGG CCTGCTGATG      14700
TCAGGGCGTC CACTGCCCCC ACACCGTCCA CAGCAGCTGT CTGAGCCCTC
             C
AATCCCCAAG CTGGCAGCTG TCATTCAGGA CCCCAACCCC TCTCAGCCCC      14800
TCTTTTCCCA CATTCATAGC CTGTAGTGCC CCCTCTAACC CCCAGTGCCA
CAGAGAAGAC GGGATTTGAA GCTGTACCCA ATTTAATTCC ATAATCAATC      14900
TATCAATTAC AGTCCGTCCA CCACCTCCCT GTGGGCTGTC CTGAGCTCTG
                ..14925]
TTGGGTTCCT GGGATGGAAT CAGTGCATCA TAAAGGGCAT TCTTTAAGCA      15000
GAGAAGGGGC CAGGCCACCC CATTCAGGAA CTGCTGCGGG AATAAAGTGC
TAACTTGCCC CCAGGCTGTC TATGGGAGAC CCTGGGCCCA GTCTGGGATG      15100
TACAGGGCTC TGGGAAGGGG GCAGTCCTGG CGGCAGAACC CGGCCTGCAG
GGGCACTTTG CTTAGAAGAG GACTCTCCTA GCGGGAGAGG CTGGGAGGGG      15200
CTGCATCAGG CCGTGGAGCT GGTTGCTGTG GTCATCAGTA TGGCTGCTTG
TTCAGGAAGC GGGAGAACAT GGTGAAGGCA GCGAGGGGCT TGTCGGTGGG      15300
AACCATGTGG CCGGCGCCCT GAGGAGCAAT GTTCGTGAGC TCCTGACCCC
ACCATTCCCT CCTCCCCATA TAACTGCTCA CTCGGGGGCA ATTCCTTCAT      15400
CCCAAACCCT TTATTCTTCC CAGAACCCTC CCCACCCCTC TCCAAAAAAA
CTTGCCCATA CAGGGGCCAG ATGGTGACCC ATGACCCAGC CTAAAAGGCA      15500
GCCAGAGGGA AAGGACGGGT GGGTCCTGCT CCTTTGCCTC CGGCCCAGTT
ATCTCTCAGC AGGCCCAGTC CCTACCTTGA TCGTGAGAAA GGCGATGTGG      15600
GAGAACTCCT TCACGAAGCC GGCAATCTGC TCCCCGCTGT CCCCGTACTT
CACTAACCAG GGCCGGCGCT GCACCTCCAT CTGCCCCACC AGGAAAGACA      15700
TCAGCCTACA GCAGCTGCAT CCTTGCTCAC AGCTACCAGC AAGACCTTAG
GGCTGGGAAT TCCTCCACAC TTGCCCTCTG TGGGCCAGAG CCAGGCAGCC      15800
AGCTGGCCAC TCCCAGGCAT ACCCGCTCCC AATCCTCCAC AGCAGCCCCT
ATCCCAGGGC CAGGAATCTC TACCTTACCT TCTGGTTGAG GGAATCCACA      15900
AACCACTCAT CCCCCATGAA ATTGCAGGCC ATGTCTACAT CTCCATTATA
TAATAGGATC TGGTATTTCT AAAGCAGGAT GGGGTAAAAA TGAGGGGTGT      16000
GGAACAAGCC CAGTCCCCAG CCCTTCCCTA GTTCAAGGCC TACCCCTCAG
```

FIGURE 1G

GAAATTCAAG GGGCCAAGCT AGATAACACG AACCAGGGAA TTTTCATGTT 16100
TTCTAACGAC TTACTGCATG TCCAGTATTC TACTAAATGT TTTATCTGTG
AAGGTAGATA TCATTATTAT CCCCATTTTA CAGATGGAGA ACCTGAGGCT 16200
CATATTTAAG AAATTTGCCC AAGATTATGT GGCTAGTACA GGGACTTGCA
CCCAGGCCAT GTGACTCCAG AGCCCATGCT ATTAACCACT GTGCTGGACA 16300
GAACCTAGTC CTAACTTGGC CCTGCCAGTG CTTCCTATAT CACCTTGTGC
AAGGTCTCCT CCCACCCTGG ACCTCTGCTG GCCCATTTTT GGCATGAATG 16400
AAGCTCTTCT CACCCTTAAG ACTCCGTGGC TGATGCAGAC CCCTAGGCCT
CCTATAGCAG GACAGCATGG GTGGGGACAA ACCCAGGTTC AAATTCTGAC 16500
CCCACCAATC ACAGGACCCT GGGCTTCGCC TTTCGGAGTT TCAGCTTCCC
TTGTCCTGAC AGGAAGTCTC AAGGGCCCCC AGCTTGCTGT GCTACTCTGG 16600
GGTCCTAAGG CTGCTGGTGA TCCAGCTGTG CTCTCCCCAC TCACCTGTGA
GCTAAGCAGC TTCAGATACT GGGAGTTCAT GCTTCGGTAG AGACGGCGGT 16700
ACTGTAAGTT TACCAGAAAG CTGCCGGGAA CATACGAGCA GGCCCCAGGA
AGACCAGGTC GGGGCGAAAA GTAGTCAACA GATCAGACCT TCTCTCCTAT 16800
CTCCAAGCTC CAACCCAGAA GGCAGACAGC TGGGTTCCAC TCCCCTTTTT
GGGGGGGGGC CACCTCTCTA GGACTGACTT GGATGTTGGG CTGGATGGGC 16900
TGGGATTCAG ACAGCCTCAA CTCACCACCA TTTAGAAAGG TACAGGGATG
CTGGGCGTGG TGGCTCACGC CTGTAATCCT AGCACTTTGG GAGACCGAGG 17000
TGAGTGGATC ACGAGGTCAG GAGTTCGAGA CCAGCCTGAC CAACATGGTG
AAACCCCGTC TCTACTAAAA AAAAAAATAC AAAAATTAGC CAGGCGTGGT 17100
GGTGCGTGCC TGTAGTCCCA GCTACTCAGA AGGCTGAGGC AGGAGAATCA
CTTGAACCCG GGAGGCGGAG GTTGCAGTGA GCCAAGATCG CAGCACTGCA 17200
CTCCAGCCTG AGTGACAGCA AGACTCCATC TCAAAAAAAA AAAGAAAAAG
AAATGTACAA GGGACGCATT TTGGTTGTTA CGAGGACTGG GATGCATCTG 17300
GCATTTAGTG TCTAGGGGCC AGCAGGTGTG GGAGTCACAA CATACAACAG
GGTAGAGTGT GAGCACTGAC CAATCAAAAT GGACCTGAGC AGCCCTGCTG 17400
GGCCCAGTGT TACCCTTTTA CTTGGGGTTG TGGGGGGGGG TGGCCACAGG
AGACTGGACA TGGAAGTTAA CAGTTTGTTA TTGTTGTTTG AGATGGAGTT 17500
TCACTTTTGT TGCCCAGGCT GGAGTGCAGT AGCATGATCT CAGCTCACTG
CAACCTCTGC CTCCTGGGTT CAAGCGATTC TCCTCCCTCA GCCTCCTGAA 17600
TAGCTGGGAT TACAGGTGCC CAACAGCACA CCTGGCTAAT TGTTGTATTT
TTAGTAGAGA TGGGGTTTCT TCATGTTGGC CAGGCTGGTC TCGAACTCCT 17700
GACCTCAGGT GATTCCCCCA CCTCAGCCTC CAAAAGTGCT AGGATTACAG
GCGTGAACCA CTGCGCCCAG TCGGAAGTAA TAGTTATTAA CCAATGTGAT 17800
GGCCGGGTGT GGGGACCCTC GCCTGTAATC CCAGCACTTT GGGAGGCCAA
GGAGGGAGGA TTGCTTGAGA CCAAGAGTTT GAGACCAACC TGGACAACAT 17900
AGTGAGACCC TGTCTCTACA AAAAAATTTT TTAAAAATTA GCCAGGAATG
GTGGCATGTG CCTGTAGACC CAGCTACTCA GTAGACTAAG GTGGGAGGAT 18000
CGCTTGAGCC TAGGAGGTTG AGGCTGTAGT GAACTGTGAA ACAACAACCA
ATGTGGCTAC GTAATGAAAT GGGGCTGGGC GGGGTGGCTC ATGCCTGTAA 18100
TCCCAGTACT TTGGGAGGCT GAAATGGGCA GATCACTTGA GCCCCAGAGT
TCAAAACCAG CCTGAGCAAT GTGGTAAAAC ACCATCTGTA CAAAAAATAC 18200
AAAAACTTGT TGGGGGTGGT GGTGCGTCTG TAGTCCCAGC TACTGGGGAG
GTGGGAGGAT TGCTTGAGCC CATGAGTTTG AGGCTGCAGT GAGCTATGAT 18300
CACACCACTG CACTCCAGCC TAGGTGACAG AGTGAGACCT TGTCTCAAAC
AAAAAAATTT TAAATTTAAA AAAGATAAAA GAATGAATTG GCTGACTCTC 18400
AATCCACTGA GGGGTCAGGG ATGGATGTTA AACATCCTCT ACAGCTCATC
TCATTGAGAA ACTCTACACA GAAGATCCCT GAGAAACCTG CCGATGCCCA 18500
CAAGCCCCAG CAACCCACCA CCCCAAGTCA CAGGTGGCCA CGGAACCTCA
CTTGCACATG TCCCATTGTG GCAGCTGCTC CGGGATGTTG AGGGCCTTCC 18600
GCACGTACGG GTTGTTGAGG TAGGTGGAAG CAGCTGTTGT GTTGGTGCAG
GGGGGGTCCA TGCGCACTTT ATCCCCTGAG CGCAGCAGTG CCTGGAAGGC 18700
ACAGACAGTG GGTCAGGGCC TCTGCTTCCC TGCCCCTCCC CCTGCCCCCA

FIGURE 1H

```
CCTCACCAGG AGGAAGCCCA CGCCCTCGCA CACCTGATGC CACATCCGCT      18800
TGAGTGGCAG GCGAGTGAAG ATGTTGCCCA AATCCTGGAC CACAACAGTG
TCCTTCTCAT ACCTGCAATG TGAGGTGGGG GAATGGGTGT GGAGCCTCCA      18900
GATGGGGCTG GGGTTGGCTC CAGGG                                 18925
```

FIGURE 1I

POLYMORPHISMS IN THE CODING SEQUENCE OF PLTP

```
ATGGCCCTCT TCGGGGCCCT CTTCCTAGCG CTGCTGGCAG GCGCACATGC
                                                    G
AGAGTTCCCA GGCTGCAAGA TCCGCGTCAC CTCCAAGGCG CTGGAGCTGG      100
TGAAGCAGGA GGGGCTGCGC TTTCTGGAGC AAGAGCTGGA GACTATCACC
ATTCCGGACC TGCGGGGCAA AGAAGGCCAC TTCTACTACA ACATCTCTGA      200
GGTGAAGGTC ACAGAGCTGC AACTGACATC TTCCGAGCTC GATTTCCAGC
CACAGCAGGA GCTGATGCTT CAAATCACCA ATGCCTCCTT GGGGCTGCGC      300
TTCCGGAGAC AGCTGCTCTA CTGGTTCTTC TATGATGGGG GCTACATCAA
CGCCTCAGCT GAGGGTGTGT CCATCCGCAC TGGTCTGGAG CTCTCCCGGG      400
                  A
ATCCCGCTGG ACGGATGAAA GTGTCCAATG TCTCCTGCCA GGCCTCTGTC
TCCAGAATGC ACGCGGCCTT CGGGGGAACC TTCAAGAAGG TGTATGATTT      500
TCTCTCCACG TTCATCACCT CAGGGATGCG CTTCCTCCTC AACCAGCAGA
TCTGCCCTGT CCTCTACCAC GCAGGGACGG TCCTGCTCAA CTCCCTCCTG      600
GACACCGTGC CTGTGCGCAG TTCTGTGGAC GAGCTTGTTG GCATTGACTA
TTCCCTCATG AAGGATCCTG TGGCTTCCAC CAGCAACCTG GACATGGACT      700
TCCGGGGGGC CTTCTTCCCC CTGACTGAGA GGAACTGGAG CCTCCCCAAC
CGGGCAGTGG AGCCCCAGCT GCAGGAGGAA GAGCGGATGG TGTATGTGGC      800
CTTCTCTGAG TTCTTCTTCG ACTCTGCCAT GGAGAGCTAC TTCCGGGCGG
                                                    A
GGGCCCTGCA GCTGTTGCTG GTGGGGGACA AGGTGCCCCA CGACCTGGAC      900
ATGCTGCTGA GGGCCACCTA CTTTGGGAGC ATTGTCCTGC TGAGCCCAGC
AGTGATTGAC TCCCCATTGA AGCTGGAGCT GCGGGTCCTG GCCCCACCGC     1000
GCTGCACCAT CAAGCCCTCT GGCACCACCA TCTCTGTCAC TGCTAGCGTC
ACCATTGCCC TGGTCCCACC AGACCAGCCT GAGGTCCAGC TGTCCAGCAT     1100
GACTATGGAC GCCCGTCTCA GCGCCAAGAT GGCTCTCCGG GGGAAGGCCC
TGCGCACGCA GCTGGACCTG CGCAGGTTCC GAATCTATTC CAACCATTCT     1200
GCACTGGAGT CGCTGGCTCT GATCCCATTA CAGGCCCCTC TGAAGACCAT
GCTGCAGATT GGGGTGATGC CCATGCTCAA TGAGCGGACC TGGCGTGGGG     1300
TGCAGATCCC ACTACCTGAG GGCATCAACT TTGTGCATGA GGTGGTGACG
AACCATGCGG GATTCCTCAC CATCGGGGCT GATCTCCACT TTGCCAAAGG     1400
GCTGCGAGAG GTGATTGAGA AGAACCGGCC TGCTGATGTC AGGGCGTCCA
CTGCCCCCAC ACCGTCCACA GCAGCTGTCT GA                        1482
 C
```

FIGURE 2

ISOFORMS OF THE PLTP PROTEIN

```
MALFGALFLA LLAGAHAEFP GCKIRVTSKA LELVKQEGLR FLEQELETIT
          R
IPDLRGKEGH FYYNISEVKV TELQLTSSEL DFQPQQELML QITNASLGLR      100
FRRQLLYWFF YDGGYINASA EGVSIRTGLE LSRDPAGRMK VSNVSCQASV
                    Y
SRMHAAFGGT FKKVYDFLST FITSGMRFLL NQQICPVLYH AGTVLLNSLL      200
DTVPVRSSVD ELVGIDYSLM KDPVASTSNL DMDFRGAFFP LTERNWSLPN
RAVEPQLQEE ERMVYVAFSE FFFDSAMESY FRAGALQLLL VGDKVPHDLD      300
MLLRATYFGS IVLLSPAVID SPLKLELRVL APPRCTIKPS GTTISVTASV
TIALVPPDQP EVQLSSMTMD ARLSAKMALR GKALRTQLDL RRFRIYSNHS      400
ALESLALIPL QAPLKTMLQI GVMPMLNERT WRGVQIPLPE GINFVHEVVT
NHAGFLTIGA DLHFAKGLRE VIEKNRPADV RASTAPTPST AAV             493
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Chew, Anne
Choi, Julie Y.
Koshy, Beena

<120> Haplotypes of the PLTP Gene

<130> MWH-0073PCT PLTP

<140> TBA
<141> 2001-03-15

<150> 60/192,127
<151> 2000-03-24

<160> 129

<170> PatentIn Ver. 2.1

<210> 1
<211> 18925
<212> DNA
<213> Homo sapien

<400> 1
gatcacttga gggcaggagt tcaagaccag cctggccaaa atggcaaaac cctgtctcca 60
ctaaaaatac aaaaattagc caggtgtggt ggcacatggc tgtagtccca gctacttggg 120
aggctgaggc aggagaatca cttgaaccca ggaggcagag gttgcagtga gccaagatca 180
caccaccgca ctgcagcctg gtgacagagc acgactgtgt ctcaaaaaaa ttaattaatt 240
aattaaataa aaaaggaaat ggaactattt ttgtgaatct gtggattata tcagaaaaaa 300
aagacacaat ggggaaagtc ctaggcaaat caggatgagt tagtcatcct tcctagatga 360
gtgtttggtg ctaaatacat gctcagcaga catgattatt gcttcccctt tctttcgtcc 420
atttggcaac aaaaaggtgg caagcaccca ctctgtgccc tgtcctaggg tccgggaacc 480
ctgtaagcag tagatggagg tgggggtggg ggtgggggcg gggatgctgt tcagagcacc 540
ttgctccaag ggttcattaa aaaatccacc agtggaccgg gcgcggtggc tcatgccttt 600
aatcccagca ctttgggagg ccgaggcggg cggatcacaa ggtcaggaga tcgagaccat 660
cctggctaac acggtgaaac cccgtctcta ctaaaaatac aaaaaaaatt agccgggcgt 720
ggcagcgtgc gcctgtagcc ccagctgctg gggaggctga ggcaggagaa tggcgtgaac 780
ccgggaggcg gggcttgcag ggagccgaga tcgcgccact gcactccagc ctgggggaca 840
gagcgagact ccgtctcaaa aaataaaaat aaaaaaaata gaaaaaacaa tccaccagcc 900
acgataaatg gcagacctcc ttctgatttc agccggtgtg gtatgttcct gggctgacag 960
cacttgtcta gtcttgcttt cccaagtggg aaaggtctct gggaccttaa ggtccccagg 1020
tggtgacaca gagacaggta gggggggccca tagcaaagcc aggcaaggag gtcccgagat 1080
gattgtgggt ggcagggaaa gaaaaaatat tccttgactt tgtgcctgga cctggttgta 1140
ataaaggccc aagaggtagt tcctatcatc gtgcacattt cgctgaagga agaaactgag 1200
ggtcagtgac ccaagtgaag tgacttgccc aagatcatgc aggaagacat ggataattgt 1260

aatttgaacc aaggtcccag caaagtggga ttgttggggc tgagtgggcc ggctcctgca 1320
tttccttccc tctccctggg cttgggtctc ccacttgtcc agacagcggc cgggcttgtc 1380
acggggctct gtgcagcctt ttccactctc ccggctgcca gcgtcccgcc ccgtcccctc 1440
ccagccccca agggaggagg ggagagctgc agagaggagg aggggtcggg gaggccggct 1500
ttataaaggc ggctggaaca accctgcccg ccagaccccg tcgcccggat cccctgagct 1560
gcccgccatc ccacgtgacc gcgccgcccc ccagctccac cgctgagtga gttggggcgc 1620
gtccctatcc cgttcccagt ccctcgccga gccgcgtggt gcactgaagc cggcgtgggg 1680
aggagagccg atgaaggaga acgctaacat gggggctcca ggcagaatct ctaatgggag 1740
agatttagga cctgagggag ccgggagacc cgggagccca cggtctggtc ggccacctcc 1800
tctcctcccc gggcgcgagg cctatgagac cgcaagccac cctccaagaa tgcgcgtgca 1860
gtcgttgccg tggcaacgtg caagcctgca tgggcccacc gctaaaggaa aggggaggtt 1920
gttgggggtg tgggggctgg tcccctggag gtccaaactc atcacgcggc ggtagcacga 1980
gggaactggg aacgtcccct ccccccgtgtc ccaaatcctc tgggagccag aacgcaagtc 2040
gcatcgcctc tctgagcctc agtttccaca tctgccatat ggaggcaaga ttattggttc 2100
caagttcccc gaagagttat ctggagggta acgatccagg ctagttccca cttatggttg 2160
gacgtccaag gaatgggttt aagcggctca gataactcct gcgaccccac cccaccccca 2220
ggcccgctcg ccatggccct cttcggggcc ctcttcctag cgctgctggc aggcgcacat 2280
gcagagttcc caggctgcaa gatccgcgtc acctccaagg cgctggagct gggtaaggcg 2340
caggggcaga cagggacgga cgggagcgga cggggttggg gtcgccccaa cagtgggcac 2400
gggagctaag taagggtttc tctgctgctt cagtgaagca ggaggggctg cgctttctgg 2460
agcaagagct ggagactatc accattccgg acctgcgggg caaagaaggc cacttctact 2520
acaacatctc tgagtaagtg gggggcgggg cctcgcgatc cggggcgggg cctcggcggt 2580
ggaggcgggg cccgagaaat ctcagggtct cgggagactg gaggctgagt gggatggggc 2640
gtggccaaat cgatgaggca agacttaaga aaccggatgg ggccgaaaag ataaaggcca 2700
agccagtcat cataggcggg accaggcaga tgggcgtggc cttgaaaatg tagaggggggc 2760
caagaaggaa gcaggaccag tggatgtgag gcgtggtcga agaacaacag aaaggtggag 2820
tcaaggaact tcccaataga tgagagtgga gtctcgggca cggggtggag ttaatcggag 2880
gggtggagaa actgaaggct ctaagtatgg tttaggatga gggaggggga ataaggaatt 2940
gagatgtgag acggagctta ggtgatggag gcaaagtgaa cggactctat taaaagaccg 3000
ggtgtagcaa ccccagggtg tgggagggga gggggtgcaa gcaaacaaga ctttttttttt 3060
ttttttttct gagagggagt cttgttctgt tgcccaggct agagtgcagt ggcgcgatct 3120
tggctcactg caacctccgc atcccgagtt caagcgattt tcctgcctca gcctcccgag 3180
tagctaggac tacaggcgcg tgccaccacg gtctgctaag ttttgtattt ttggtagaga 3240
cgggtttcac tatgttggcc agactggtct cgaactcctg acctcaagtg atccgccctc 3300
ctcggcctcc taaagtgctg ggattacagg catgagccac caccgcacct ggcaaacaaa 3360
caaacaagac ttttaagcga tggggtgaag ctcatttttа acacctggat caggaagacc 3420
tcctctctca gccacccaag ttcaactttg agccacccaa gtctaagttg ggcacttgag 3480
ggacaatgga agaattcatt ccacccactc ccgcttcccc aggcctggac ttgaaagggg 3540
agcagacaaa tttcctgtcg ttgggggaag ttccctcttc ttggccctgg atctgaccct 3600
gaggcctcct gtagggtgaa ggtcacagag ctgcaactga catcttccga gctcgatttc 3660
cagccacagc aggagctgat gcttcaaatc accaatgcct ccttggggct gcgcttccgg 3720
agacagctgc tctactggtt cttgtaagga cccagcaccc tcaggggagt ggagagtggg 3780
tcaggagggc tggaggtgct gtggggctgt gggcagcccc cttagccttc ctgagtttcc 3840
gtttcttcat ctgtgtaaca gaaacaggca tatgcatttc atacgagttg ataggaaggt 3900
ttggtaggat gtgtggaaag aggtgaaacc ttattaagag gtatggtgat gcatgggggc 3960
actgaagact caattgagct cctacctgga gtgaatatta acccccctgg cagctatgat 4020
gggggctaca tcaacgcctc agctgagggt gtgtccatcc gcactggtct ggagctctcc 4080
cgggatcccg ctggacggat gaaagtgtcc aatgtctcct gccaggcctc tgtctccaga 4140

atgcacgcgg ccttcggggg aaccttcaag taagccccag ccccaacccc agctcaccct 4200
tttcgagccc cacagctcat ctgcctagcg caatgcaaag tgcataggcc tgggaatcag 4260
accaactggg tccaaagctc agctctgtca cttcctggct gtgtgaacgt ggccaagtca 4320
ctgaagcttg gtgcctcaat atcctcattt tataagatcc tattttataa gatcgttgtc 4380
aggattattt gtgttggtat atgtaaaaaa tatttagaac aaggctggct gtggtggctc 4440
atgcctgaaa tcccagcaat ttgggaggtc tgggcgggcg gatcacaagg ttaacagatc 4500
gagatcatcc tggccaacat ggtgaaaccc catctctact aaaaatacaa aaattagctg 4560
ggcgtggtgg cacgtgcctg tagtcccagc tacgtgggag gctgaggcag gagaatcgct 4620
tgaacccggg aggcggaggt tgcagtgagc tgagatcgtg ccactgcact ccagcctggt 4680
gacagagcaa gattccatct caaaaaaaaa aaaaaaaaaa aatttagaac agcacctggc 4740
atatgataaa tgtatattta agtgctggcc acttagataa gagtaactgt gaggactatc 4800
tgatatgcac atttggccat gtactggatt ctcaataatt ggtatctcta aataataata 4860
acaaaaataa taatttttt tttgagaggg agtctcactg tgtcacccag cctggagtgc 4920
agtggcgcca tctcggttta ctgcaacctc cacctcctgg gttcaaacga ttctcctgcc 4980
tcagcctccc aagtagctgg gactacaggc atgcaccacc atgcccaact actttttttg 5040
tatttttagt agagacaggg tttcgccata ttggccacgc tggtctcaaa ctcctgatct 5100
caggtggtct gtccgcctcg gcctaccaaa gtgctgggat tacaggcctg agcaactgtg 5160
cccggcctat ttttttattt ttgctgctga gaaggggata ttgttcttgt tctggaagag 5220
gatgggggag agaagatgat ggattataac ctggtgttgg tacactgtca gtattattta 5280
ttcatcaaac aaagtgtatg ccagtccctg aggatgcagg aaggagctca gagcctatct 5340
ggagcaaaaa acatttgata tcagggcctg gagagaagga ctaagaaaat gctttgggga 5400
ttcagaggag gggaatctgg gaagactgcc tggaagagga agcatctgag ctcagatggg 5460
aaaatacaga gtaggaatgc agagggcgga agggagggca tcagtaagcc gatggatgtg 5520
gggatgctca gagtgggttt gaggcaatgt gggtggattc atttgactga tgggaccaga 5580
gagtaggtca ggaggtcttg agaattagac tggagaggga atttgggcct ggttcttgga 5640
aaacctacgt ggaggagctg ttattcaggc tgaagctgag gactcctact gccactattt 5700
ccctagtcac tgatttctcc tggacctgga ctggggtgga ggcaggttct gggctcatcc 5760
ggcctctcct ctcctcacag gaaggtgtat gattttctct ccacgttcat cacctcaggg 5820
atgcgcttcc tcctcaacca gcaggtgtgg gcagcgacag gtcgcagggt ggcaagggtg 5880
ggcatgctct cactttgaga aggccctgac tctggctccc acctcgcaga tctgccctgt 5940
cctctaccac gcagggacgg tcctgctcaa ctccctcctg gacaccgtgc ctggtgagtt 6000
ggtggcgggt gagtctgggt gtgcagctgt catgcagcac ctcagagcag gccccttccg 6060
agccctgctg ttgaacagtc ctgggttcaa atgtggcccc tggagctgac ttgttgtgcg 6120
attttgggtg aagtgctcat ttctgtctgg gcaaagtgct tgtttcctca cttggctggc 6180
agtgcctcct cctggggttg ctgtgaggat tatctgagag aatagttgtc aaggtcctca 6240
atatgtgcat gctgatcact ggcatcccat ttgtcttcat cacctagcat ctaataggtg 6300
ctcaattaac atgaattccc ttttttcttt tctttcttt ttttttttt ttgcacggag 6360
tcttacagtc tcactctgtt gccaggctgg agtccagtgg cgcgatctca gctcactgca 6420
atctccacct cccgggttca agcaattctc ctgcctcagc atctcgagta gctgggacta 6480
caggcgtgca ccaccatgcc caactagttt tttgtgtttt agtagagacg gggtttcacc 6540
atgttggccg ggatggtctc gatctcctga cctcctgatc tgcccacctt ggcctcccaa 6600
agtgctggga ttataggcat gagccactgt gcctggccat gaattccctt tttctaagcc 6660
atgacctctg tctactgtgg ggcaaaggta ttttcccatc acacctaaca cacaggcatg 6720
ggaagcgaag gaggaggtag gagaaaccaa gggaaaggaa cctcatgcca tctttttttt 6780
ttttttttt ttttgagatg gagtctcgct ctgtcaccca ggctggagtg cagtggtgca 6840
atcctggctc cctgcaacct ccacctccca ggatcaagcg attctcctgc ctcagcctcc 6900
tgagtagctg ggattacagg cgcccaccac catgcccaac taatttttgt atttttagta 6960
gagacggggt ttcaccgtgg tggaccaggc tggtctcgaa cacctgacct caggtgatcc 7020

gcccgccttg acctccctaa gtgctgggat tacaggtgtg agccaccatg cgcggccccc 7080
ccatgccatc ttgtttagtt attttttttga aacctctata gtggtagtaa taataataac 7140
taacattaat gagcacctaa ctacacacca ggccctaagt gctttccaca taaaactaat 7200
ttgaccctca tgacaaccct gtgaaggatg tatcctcatt ttatgggtgg gcctctgagc 7260
ctggagaggt taagcacctt atcccacatc acacagccgt gagtgacttg agccccttcc 7320
tgcagtgcgc agttctgtgg acgagcttgt tggcattgac tattccctca tgaaggatcc 7380
tgtggcttcc accagcaacc tggacatgga cttccgggtg agctgcttgg gctggtgtat 7440
gacctctgac ttcctaacaa gacctctttc tctgctaagt tgcaagattt tacttcttaa 7500
agcacagctc tgatgaggcc actcccatta ctgagaatgc tatggctccc tgttactaca 7560
gaattaggtt cagcctcctg ggtctggcat ttgagcccct gttaatctta cctggcccca 7620
ccttttatat ctcatcctgt ttctttgcag ctgctctttc ccaaattgga ccctggcttc 7680
ccacctcagg gcctttgctc atgcacggcc tttccccaac tcccccgaag gagctggaca 7740
taaatttcta tacaaataat ctcatttcat tctctcacaa ccctgtgagt caggtactaa 7800
catcatccta attttaggaa tgtggagact tactcataga ggtgacgtca cttgcccaag 7860
gccacatagc cagtgtgtag ccactacctt tggtctccca gaggagtata gcagtttcag 7920
agtgcagact cgggagctag aatcagtcgg gtgcagtggc tcatgcctgt aatcccaata 7980
ctttgggagg cccaggcgag aggatcgctt gagcccagga gtttgagacc agcttgggca 8040
tcatagtgag acctccttct ctacaaagag aaaaagaatg cctgaagttc aaatcctagc 8100
tctgccactg actagctgca cggccctaag cataagtttt cagttggcca gtttcctaac 8160
atgtaaatgg aaataatcta gtaagaaaac acacgttgtt atgcagatta aagaagtcga 8220
tctatgtgaa gagcttagag tacttggcaa agaagcatga gtatgcacaa tatatgtaaa 8280
cagtagctgt tgttactgcc tctagctttc aaggcccact cagatgtcac ctcttccctg 8340
aaacctttcc tgaccctccc tcctccagtc tgacgggctc gccccctggt agcaggtgtc 8400
ctccttccaa tgccttttaa gagatgtgct tgtgaatcct ccaagaacag cagctgggtt 8460
cccaccatcc ccccacagtg cctagcacgg ggctgggaga cacctgctgt cagtccaggt 8520
gccaggccag tctgccttga ccctgactgt gaatgcccca ccccaggggg ccttcttccc 8580
cctgactgag aggaactgga gcctccccaa ccgggcagtg gagccccagc tgcaggagga 8640
agagcggatg gtgtatgtgg ccttctctga gttcttcttc gactctgcca tggagagcta 8700
cttccgggcg ggggccctgc agctgttgct ggtgggggac aaggtatgtc atggcctgtt 8760
tgtgggatgg gcaagagaag gtctgtgaca gagctcactc cctcactcct gattcccctg 8820
ttcaggtgcc ccacgacctg gacatgctgc tgagggccac ctactttggg agcattgtcc 8880
tgctggtgag tgctggcggg ggcagggata gggccacctg cacaccagtg agaccaagga 8940
ggggtgcagg tggggccccc agtggcagcc acgcagacag ggccctggcc tcttgtctgt 9000
ggccaatccc cctcgccttg tcatggctgt tcctgtctgt gaagcggtct gctccttcct 9060
gctccttcct gccccatctc cttcgcagag ctgctgtgag gattagtgag aacccaagct 9120
ttgcaagtgt gagggcttta ataataattc acatatactt ttgtttattt gtgtattggg 9180
gaggagagta agagttacac attttattta ttttttttt tgagacaaga gtctcacttt 9240
gtcacccaag ttggaatgca gtggcacaat ctaggctcac tgcaacctct gcctcccagg 9300
gctcaagcca tcctcctgct tcagcctctt gagtagctgg aattacaggc ttgtgccacc 9360
acaccccgct aatttttata ttttgttgag acagagtttc attatgttgc tcaggctgtc 9420
cttgaactcc tgggctcaag tgatcaccca tttttggcct cccaaagtgc taggattaca 9480
ggtgtgagcc accgcaccta gcagaattac aaattttaaa agaaataatt cagtcattat 9540
catcatccaa tttatggcag ggcagggtct ctattgtcag actgcctggg cttcaagcct 9600
agttctgaaa ctgagtaact acataacctc tacaagttac ttaacctatc tgtgcctcaa 9660
tttccccatc tcaaaaatgg ggatatggct gggcatgatg gctcacgcct atactctgag 9720
aactttgaga ggtcaaggta ggaaaattgc ttgagcccag gagtttgaga ccagcctggg 9780
caacacagtg agacactgtc tctattgaaa aataaatttt tagtcagatc ttttttttc 9840
ggtgggggga caaagtctca ctctgtcacc caggctggag tgcagtggcg caatcttggc 9900

tcactgcaat tgaacctccc gggttcaagc aattctctgc catgacgact cagcctgaaa 9960
aataaatttg taaaaaataa caacaacaaa attggcagta ataatatcgg cctcacatgg 10020
ttattatgag gtttaaagga accaacacat atacagtgtt tagaatcatg tttggtgaag 10080
attaggcagt tgcatcttat agtattcctt cactgcctat agaagaaggc atagcagttt 10140
agagaaagaa gatatggcca ggcacggtgg ctcacgcctg taatcccaac actgtgggag 10200
gctgaagccg gcagatcact tgagcccagg agtttgagac cagcatgggc aacatggcca 10260
aaccccgtct ctacgaaaac acaaaagcaa aaaccaaaac aaaaaacaaa tacaaaaata 10320
agccaggcgt gatggcatga acctgtggtc tcagctgttc aggaggctga tgtggcagga 10380
taatttgagc tgaggaggtt gaggctgcag tgagccgtga tcatgccact gcactccagc 10440
ctgggcaaca gagaccctgt ctcaaaaaaa aaggagagag agagcaaaaa gatatggcaa 10500
ggggtttggg tgtggtggct catgcctgta atcccagcac tttgggaggc cgaggtgggc 10560
gaatcactag cctggccaat atggcgaaac cccgtttcta ttaaaaatac aaaacttagc 10620
cgggcatggc ggcaggtgcc tgtaatctca gctacttggg aggctgaggc gggagaatcg 10680
cctgaacctg ggagacagag ttgcagtgag ccgagattgt gccactgcac tccagcctgg 10740
gcaacaaagt gagactccat atcaaaaaca gaagacatgg caagggctga gatactaatg 10800
caaaaaggac atccacaaag agtctggatt ggagatagac ctgaatattg agatgtattt 10860
aggaagcaga aagagaagaa cacattcctg tttctttgtg tggcaaaggc atgaaggcta 10920
gatgcagatg tgagagtttt ggaagccaga gctccctctg tatttgcaat atgagagtcc 10980
ccagaggaca gaagctgcac ctgccttcct tccccttctt ccctgggggc tccaaaggaa 11040
atggcaggga cttgacccgc ccctgctgct catcccccag agcccagcag tgattgactc 11100
cccattgaag ctggagctgc gggtcctggc cccaccgcgc tgcaccatca agccctctgg 11160
caccaccatc tctgtcactg ctagcgtcac cattgccctg gtcccaccag accagcctga 11220
ggtccagctg tccagcatga ctatggtacg ggtcccagag tgggggctgc tggctggggg 11280
tcagggaaaa atgggctgtg ggttactgac cttctgggtc agtaacatcc tcctccccat 11340
gttcctgcag gacgcccgtc tcagcgccaa gatggctctc cgggggaagg ccctgcgcac 11400
gcagctggac ctgcgcaggt aggcaggcgc acttccctgc acctcagtga ttggaggaca 11460
atcacacctg tccacaggga tgttgcgctt aaagtacagt gatacctgta agccactgag 11520
agctggggtc tgtctttgtt aatgcatgcc aaagtgcttt ggtaaagtct gactcctctg 11580
cattctcaca tgtcccagag gcaggcaggg caggtattat gagagatgaa caaatgggtc 11640
cagcaatgtt aaataacatg cccacagtca cacagcaaag atcagaacat gggttttgac 11700
aaaaacctca gcacaacaca gctttgtgac cttgaggcca ttttatcttt tttaaaaatg 11760
ttattattct ttttttttt tttttgagaa ggagtttcgc tcttgttgcc caggctggag 11820
tgcaatagcg tggtcttgac tcactgccac ctctgcctcc cgggtttagg caattctgcc 11880
tcagcttcct gagtagctgg gattacaggt gtgcgccacc acgcccggct aatttttgca 11940
tttttagtag agacagggtt tcaccatatt ggtcaggctg gtctcaaact cctgacctca 12000
ggtgattcac ctgccccggc ctcccaaagt gctgggatta taggtatgag ccactgtgct 12060
tggccttatt ttattgtatt tatttatttt tgagacagag tctcactctg ttgcccaagc 12120
tggagtgcag tggtgcaatc atagctcact gcagccttga cctcctgggc tcaagtgatc 12180
gtccagcctc agcctcccga gtaactggga ctgcatgcat gcactaccat gcctgactat 12240
tttatttttt gtagagatgg ggtctcactt tgttgtccag gctggtctcc aattctcggg 12300
ctcaagtgat cgtcccacct tagcctccca aagtgctagg attacaggcg tgagccaccg 12360
cacccagccc attttgtctt tctaggcctc attttctcca tgtgtttaat acaggcaaaa 12420
gtagcatctg ccttctaggg ctgtgctatc caatatggta gatactagcc acatgtgact 12480
atttacattt aaattaatta taattaaata gtttttcagt tgcactaagc cacatttcat 12540
tagccacatg tggctagtga ctaccatagt agacggtgca gatacaaaac gtttcacaga 12600
aaatctgact ggacagtgct gttctaaggt taaaggagat aatgtaggtt aggaccctag 12660
gttctagagg cagcttctga cactttccag ctgtgtgact tcaggcaaat tacttaacct 12720
tgcagggcct cattttcctc atctataaat tagtgcaaat cctaatagtg cccactcctt 12780

gggcttgttt tgaggactaa accagttggt actgtatttt aaaacagttt actaatatta 12840
acttttaaa cacgttggct gggtgcagtg gctcacacct gtaatctcaa cactttggga 12900
gaccaaggcg ggtagatcgc ttgaggccag gagttcgata ccagcctgag ctacaaagtg 12960
agatcctgtc tctacaaaaa aatttaaaag ttagttggtc atagtgatgc acgcctgtgg 13020
tcccagctac ttggaagcct gaggtgggag gctgacttga gccctggagg tcaaagccac 13080
agtgagtaaa attacaccac tgccctccag cctgggcaac acagtgagac gctgtctcaa 13140
aaataaataa aaataaaaaa agaaaataaa tatgcttagt gtctgacagg aactaagcaa 13200
tgtgaagttt tgttacataa aataaataac acctgcaaag ctaccagcag aagtctggca 13260
cattagaagt gctcaataaa ggccgggcgt ggtggctcac gcctgtaatc ccagcacttt 13320
gggaggccga gacgggcgga tcacaaggtc aggaaatccg agaccatcct ggctaacacg 13380
gtgaaacccc gtctctacta aaaacacaaa aaattagctg ggcttggtgg cgggcaccta 13440
tagtcccagt tactcgggag gctgagccag gagaatggcg tgaacccggg agacggagca 13500
gtgagccgag atcgcgccac tgcactccag cgtgggcgac acagcgagac tctgtctcaa 13560
aaaaaaaaaa aaaaaaaaaa aaaaaagagg tgctcaataa ataaaaagta gcagtcttgg 13620
agaaaaggag acccaagtgg cagcaatgac tgccccgagg atacctggaa agttgtagca 13680
cttggaaagg tgcctaggtg ggcatctgag cctgcaactc agacacctct ggcaggtagg 13740
cctgggtggc ctgaggaagg ggagcttcca ggcttcctgc ccacagggct cctttctttc 13800
ccgcaggttc cgaatctatt ccaaccattc tgcactggag tcgctggctg tgagtgggga 13860
agggtgggtg ggctgggctg ggttcgtgag atcctgaccc tcaccatccc cccctcctca 13920
tataactgct cacctgcggg cagctaaagg caaacccaac cctgactcac tgccctgttt 13980
ccgccctcgc ccctcccagc tgatcccatt acaggccccc ctgaagacca tgctgcagat 14040
tggggtgatg cccatgctca atggtaaggc tggggtgtga ggatggagga agaaaggagg 14100
ggtgaactgg gcgggcccag actgagcggg gtgctcccac ccacagagcg gacctggcgt 14160
ggggtgcaga tcccactacc tgagggcatc aactttgtgc atgaggtggt gacgaaccat 14220
gcggtgagtg ggggcaggat ggggaaggag ggagggcctt cccctctctc tatagctccc 14280
cagtgccctc cattacctga ccccagctcc cccattacca gtcgcccctc ccattttaat 14340
cttagactcc taaaagccgt tgcacacact gtgacttctt ggagcactct tctcccatcc 14400
caatacctac taattgttca ggtcttagct cagaggtcac ctcctccaag aagccctcct 14460
tgacacccat agtctacact cctacagcct cattttaaca ccttgacatg gaattatcca 14520
ttagattgag ttcctgaaag acaaacactc tgtctcttga tctctaaact ctggcctcaa 14580
taactgaatg aataaatgag gctcccctct ccccaacaca gggattcctc accatcgggg 14640
ctgatctcca ctttgccaaa gggctgcgag aggtgattga gaagaaccgg cctgctgatg 14700
tcagggcgtc cactgccccc acaccgtcca cagcagctgt ctgagccctc aatccccaag 14760
ctggcagctg tcattcagga ccccaacccc tctcagcccc tcttttccca cattcatagc 14820
ctgtagtgcc ccctctaacc cccagtgcca cagagaagac gggatttgaa gctgtaccca 14880
atttaattcc ataatcaatc tatcaattac agtccgtcca ccacctccct gtgggctgtc 14940
ctgagctctg ttgggttcct gggatggaat cagtgcatca taaagggcat tctttaagca 15000
gagaaggggc caggccaccc cattcaggaa ctgctgcggg aataaagtgc taacttgccc 15060
ccaggctgtc tatgggagac cctgggccca gtctgggatg tacagggctc tgggaagggg 15120
gcagtcctgg cggcagaacc cggcctgcag gggcactttg cttagaagag gactctccta 15180
gcgggagagg ctgggagggg ctgcatcagg ccgtggagct ggttgctgtg gtcatcagta 15240
tggctgcttg ttcaggaagc gggagaacat ggtgaaggca gcgaggggct tgtcggtggg 15300
aaccatgtgg ccggcgccct gaggagcaat gttcgtgagc tcctgacccc accattccct 15360
cctccccata taactgctca ctcgggggca attccttcat cccaaaccct ttattcttcc 15420
cagaaccctc cccacccctc tccaaaaaaa cttgcccata caggggccag atggtgaccc 15480
atgacccagc ctaaaaggca gccagaggga aaggacgggt gggtcctgct cctttgcctc 15540
cggcccagtt atctctcagc aggcccagtc cctaccttga tcgtgagaaa ggcgatgtgg 15600
gagaactcct tcacgaagcc ggcaatctgc tccccgctgt ccccgtactt cactaaccag 15660

ggccggcgct gcacctccat ctgccccacc aggaaagaca tcagcctaca gcagctgcat 15720
ccttgctcac agctaccagc aagaccttag ggctgggaat tcctccacac ttgccctctg 15780
tgggccagag ccaggcagcc agctggccac tcccaggcat acccgctccc aatcctccac 15840
agcagcccct atcccagggc caggaatctc taccttacct tctggttgag ggaatccaca 15900
aaccactcat cccccatgaa attgcaggcc atgtctacat ctccattata taataggatc 15960
tggtatttct aaagcaggat ggggtaaaaa tgaggggtgt ggaacaagcc cagtccccag 16020
cccttcccta gttcaaggcc taccccctcag gaaattcaag gggccaagct agataacacg 16080
aaccagggaa ttttcatgtt ttctaacgac ttactgcatg tccagtattc tactaaatgt 16140
tttatctgtg aaggtagata tcattattat ccccatttta cagatggaga acctgaggct 16200
catatttaag aaatttgccc aagattatgt ggctagtaca gggacttgca cccaggccat 16260
gtgactccag agcccatgct attaaccact gtgctggaca gaacctagtc ctaacttggc 16320
cctgccagtg cttcctatat caccttgtgc aaggtctcct cccaccctgg acctctgctg 16380
gcccattttt ggcatgaatg aagctcttct cacccttaag actccgtggc tgatgcagac 16440
ccctaggcct cctatagcag gacagcatgg gtggggacaa acccaggttc aaattctgac 16500
cccaccaatc acaggaccct gggcttcgcc tttcggagtt tcagcttccc ttgtcctgac 16560
aggaagtctc aagggccccc agcttgctgt gctactctgg ggtcctaagg ctgctggtga 16620
tccagctgtg ctctccccac tcacctgtga gctaagcagc ttcagatact gggagttcat 16680
gcttcggtag agacggcggt actgtaagtt taccagaaag ctgccgggaa catacgagca 16740
ggccccagga agaccaggtc ggggcgaaaa gtagtcaaca gatcagacct tctctcctat 16800
ctccaagctc caacccagaa ggcagacagc tgggttccac tccccttttt gggggggggc 16860
cacctctcta ggactgactt ggatgttggg ctggatgggc tgggattcag acagcctcaa 16920
ctcaccacca tttagaaagg tacagggatg ctgggcgtgg tggctcacgc ctgtaatcct 16980
agcactttgg gagaccgagg tgagtggatc acgaggtcag gagttcgaga ccagcctgac 17040
caacatggtg aaaccccgtc tctactaaaa aaaaaaatac aaaaattagc caggcgtggt 17100
ggtgcgtgcc tgtagtccca gctactcaga aggctgaggc aggagaatca cttgaacccg 17160
ggaggcggag gttgcagtga gccaagatcg cagcactgca ctccagcctg agtgacagca 17220
agactccatc tcaaaaaaaa aaagaaaaag aaatgtacaa gggacgcatt ttggttgtta 17280
cgaggactgg gatgcatctg gcatttagtg tctaggggcc agcaggtgtg ggagtcacaa 17340
catacaacag ggtagagtgt gagcactgac caatcaaaat ggacctgagc agccctgctg 17400
ggcccagtgt tacccttta cttggggttg tgggggggg tggccacagg agactggaca 17460
tggaagttaa cagtttgtta ttgttgtttg agatggagtt tcacttttgt tgcccaggct 17520
ggagtgcagt agcatgatct cagctcactg caacctctgc ctcctgggtt caagcgattc 17580
tcctccctca gcctcctgaa tagctgggat tacaggtgcc caacagcaca cctggctaat 17640
tgttgtattt ttagtagaga tggggtttct tcatgttggc caggctggtc tcgaactcct 17700
gacctcaggt gattcccca cctcagcctc caaaagtgct aggattacag gcgtgaacca 17760
ctgcgcccag tcggaagtaa tagttattaa ccaatgtgat ggccgggtgt ggggaccctc 17820
gcctgtaatc ccagcacttt gggaggccaa ggagggagga ttgcttgaga ccaagagttt 17880
gagaccaacc tggacaacat agtgagaccc tgtctctaca aaaaaatttt ttaaaaatta 17940
gccaggaatg gtggcatgtg cctgtagacc cagctactca gtagactaag gtgggaggat 18000
cgcttgagcc taggaggttg aggctgtagt gaactgtgaa acaacaacca atgtggctac 18060
gtaatgaaat ggggctgggc ggggtggctc atgcctgtaa tcccagtact ttgggaggct 18120
gaaatgggca gatcacttga gccccagagt tcaaaaccag cctgagcaat gtggtaaaac 18180
accatctgta caaaaaatac aaaaacttgt tgggggtggt ggtgcgtctg tagtcccagc 18240
tactggggag gtgggaggat tgcttgagcc catgagtttg aggctgcagt gagctatgat 18300
cacaccactg cactccagcc taggtgacag agtgagacct tgtctcaaac aaaaaaattt 18360
taaatttaaa aaagataaaa gaatgaattg gctgactctc aatccactga ggggtcaggg 18420
atggatgtta aacatcctct acagctcatc tcattgagaa actctacaca gaagatccct 18480
gagaaacctg ccgatgccca caagccccag caacccacca ccccaagtca caggtggcca 18540

cggaacctca cttgcacatg tcccattgtg gcagctgctc cgggatgttg agggccttcc 18600
gcacgtacgg gttgttgagg taggtggaag cagctgttgt gttggtgcag ggggggtcca 18660
tgcgcacttt atcccctgag cgcagcagtg cctggaaggc acagacagtg ggtcagggcc 18720
tctgcttccc tgcccctccc cctgccccca cctcaccagg aggaagccca cgccctcgca 18780
cacctgatgc cacatccgct tgagtggcag gcgagtgaag atgttgccca aatcctggac 18840
cacaacagtg tccttctcat acctgcaatg tgaggtgggg gaatgggtgt ggagcctcca 18900
gatggggctg gggttggctc caggg 18925

<210> 2
<211> 1482
<212> DNA
<213> Homo sapien

<400> 2
atggccctct tcggggccct cttcctagcg ctgctggcag gcgcacatgc agagttccca 60
ggctgcaaga tccgcgtcac ctccaaggcg ctggagctgg tgaagcagga ggggctgcgc 120
tttctggagc aagagctgga gactatcacc attccggacc tgcggggcaa agaaggccac 180
ttctactaca acatctctga ggtgaaggtc acagagctgc aactgacatc ttccgagctc 240
gatttccagc cacagcagga gctgatgctt caaatcacca atgcctcctt ggggctgcgc 300
ttccggagac agctgctcta ctggttcttc tatgatgggg gctacatcaa cgcctcagct 360
gagggtgtgt ccatccgcac tggtctggag ctctcccggg atcccgctgg acggatgaaa 420
gtgtccaatg tctcctgcca ggcctctgtc tccagaatgc acgcggcctt cgggggaacc 480
ttcaagaagg tgtatgattt tctctccacg ttcatcacct cagggatgcg cttcctcctc 540
aaccagcaga tctgccctgt cctctaccac gcagggacgg tcctgctcaa ctccctcctg 600
gacaccgtgc ctgtgcgcag ttctgtggac gagcttgttg gcattgacta ttccctcatg 660
aaggatcctg tggcttccac cagcaacctg gacatggact tccgggggggc cttcttcccc 720
ctgactgaga ggaactggag cctccccaac cgggcagtgg agccccagct gcaggaggaa 780
gagcggatgg tgtatgtggc cttctctgag ttcttcttcg actctgccat ggagagctac 840
ttccgggcgg gggccctgca gctgttgctg gtgggggaca aggtgcccca cgacctggac 900
atgctgctga gggccaccta ctttgggagc attgtcctgc tgagcccagc agtgattgac 960
tccccattga agctggagct gcgggtcctg gccccaccgc gctgcaccat caagccctct 1020
ggcaccacca tctctgtcac tgctagcgtc accattgccc tggtcccacc agaccagcct 1080
gaggtccagc tgtccagcat gactatggac gcccgtctca gcgccaagat ggctctccgg 1140
gggaaggccc tgcgcacgca gctggacctg cgcaggttcc gaatctattc caaccattct 1200
gcactggagt cgctggctct gatcccatta caggcccctc tgaagaccat gctgcagatt 1260
ggggtgatgc ccatgctcaa tgagcggacc tggcgtgggg tgcagatccc actacctgag 1320
ggcatcaact ttgtgcatga ggtggtgacg aaccatgcgg gattcctcac catcggggct 1380
gatctccact ttgccaaagg gctgcgagag gtgattgaga agaaccggcc tgctgatgtc 1440
agggcgtcca ctgcccccac accgtccaca gcagctgtct ga 1482

<210> 3
<211> 493
<212> PRT
<213> Homo sapien

<400> 3

```
Met Ala Leu Phe Gly Ala Leu Phe Leu Ala Leu Leu Ala Gly Ala His
 1               5                  10                  15

Ala Glu Phe Pro Gly Cys Lys Ile Arg Val Thr Ser Lys Ala Leu Glu
            20                  25                  30

Leu Val Lys Gln Glu Gly Leu Arg Phe Leu Glu Gln Glu Leu Glu Thr
        35                  40                  45

Ile Thr Ile Pro Asp Leu Arg Gly Lys Glu Gly His Phe Tyr Tyr Asn
    50                  55                  60

Ile Ser Glu Val Lys Val Thr Glu Leu Gln Leu Thr Ser Ser Glu Leu
65                  70                  75                  80

Asp Phe Gln Pro Gln Gln Glu Leu Met Leu Gln Ile Thr Asn Ala Ser
                85                  90                  95

Leu Gly Leu Arg Phe Arg Arg Gln Leu Leu Tyr Trp Phe Phe Tyr Asp
            100                 105                 110

Gly Gly Tyr Ile Asn Ala Ser Ala Glu Gly Val Ser Ile Arg Thr Gly
        115                 120                 125

Leu Glu Leu Ser Arg Asp Pro Ala Gly Arg Met Lys Val Ser Asn Val
    130                 135                 140

Ser Cys Gln Ala Ser Val Ser Arg Met His Ala Ala Phe Gly Gly Thr
145                 150                 155                 160

Phe Lys Lys Val Tyr Asp Phe Leu Ser Thr Phe Ile Thr Ser Gly Met
                165                 170                 175

Arg Phe Leu Leu Asn Gln Gln Ile Cys Pro Val Leu Tyr His Ala Gly
            180                 185                 190

Thr Val Leu Leu Asn Ser Leu Leu Asp Thr Val Pro Val Arg Ser Ser
        195                 200                 205

Val Asp Glu Leu Val Gly Ile Asp Tyr Ser Leu Met Lys Asp Pro Val
    210                 215                 220

Ala Ser Thr Ser Asn Leu Asp Met Asp Phe Arg Gly Ala Phe Phe Pro
225                 230                 235                 240

Leu Thr Glu Arg Asn Trp Ser Leu Pro Asn Arg Ala Val Glu Pro Gln
                245                 250                 255
```

```
Leu Gln Glu Glu Glu Arg Met Val Tyr Val Ala Phe Ser Glu Phe Phe
            260                 265                 270

Phe Asp Ser Ala Met Glu Ser Tyr Phe Arg Ala Gly Ala Leu Gln Leu
        275                 280                 285

Leu Leu Val Gly Asp Lys Val Pro His Asp Leu Asp Met Leu Leu Arg
    290                 295                 300

Ala Thr Tyr Phe Gly Ser Ile Val Leu Leu Ser Pro Ala Val Ile Asp
305                 310                 315                 320

Ser Pro Leu Lys Leu Glu Leu Arg Val Leu Ala Pro Pro Arg Cys Thr
                325                 330                 335

Ile Lys Pro Ser Gly Thr Thr Ile Ser Val Thr Ala Ser Val Thr Ile
            340                 345                 350

Ala Leu Val Pro Pro Asp Gln Pro Glu Val Gln Leu Ser Ser Met Thr
        355                 360                 365

Met Asp Ala Arg Leu Ser Ala Lys Met Ala Leu Arg Gly Lys Ala Leu
    370                 375                 380

Arg Thr Gln Leu Asp Leu Arg Arg Phe Arg Ile Tyr Ser Asn His Ser
385                 390                 395                 400

Ala Leu Glu Ser Leu Ala Leu Ile Pro Leu Gln Ala Pro Leu Lys Thr
                405                 410                 415

Met Leu Gln Ile Gly Val Met Pro Met Leu Asn Glu Arg Thr Trp Arg
            420                 425                 430

Gly Val Gln Ile Pro Leu Pro Glu Gly Ile Asn Phe Val His Glu Val
        435                 440                 445

Val Thr Asn His Ala Gly Phe Leu Thr Ile Gly Ala Asp Leu His Phe
    450                 455                 460

Ala Lys Gly Leu Arg Glu Val Ile Glu Lys Asn Arg Pro Ala Asp Val
465                 470                 475                 480

Arg Ala Ser Thr Ala Pro Thr Pro Ser Thr Ala Ala Val
                485                 490
```

<210> 4
<211> 15

<212> DNA
<213> Homo sapien

<400> 4
tgtgcctrga cctgg 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
ctatcccstt cccag 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
tgagggascc gggag 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
cgcatcgyct ctctg 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapien

<400> 8
ggctagtycc cactt 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapien

<400> 9
ggcgcacrtg cagag 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
gagtgggrtg gggcg 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
gagtttcmgt ttctt 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapien

<400> 12
tttcatayga gttga 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13
gataggargg tttgg 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
ggtgtgtmca tccgc 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
gtgtgaaygt ggcca 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
taacaagrcc tcttt 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
tccgggcrgg ggccc 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
aggtgggkcc cccag 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
gtggcaamgg catga 15

<210> 20
<211> 15

<212> DNA
<213> Homo sapien

<400> 20
caaaggarat ggcag 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
cttgaccygc ccctg 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
aggtaggsag gcgca 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapien

<400> 23
gatgttgmgc ttaaa 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapien

<400> 24
atcccccsct cctca 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapien

<400> 25
ctcacctkcg ggcag 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
ctgtttcygc cctcg 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
gtctcttrat ctcta 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapien

<400> 28
cgtccacygc cccca 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29
tgactttgtg cctrg 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
ttacaaccag gtcya 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapien

<400> 31
gcgtccctat cccst 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapien

<400> 32
gagggactgg gaasg 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapien

<400> 33
aggacctgag ggasc 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapien

<400> 34
ccgggtctcc cggst 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapien

<400> 35
gcaagtcgca tcgyc 15

<210> 36
<211> 15

<212> DNA
<213> Homo sapien

<400> 36
gaggctcaga gagrc 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapien

<400> 37
gatccaggct agtyc 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapien

<400> 38
aaccataagt gggra 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapien

<400> 39
ctggcaggcg cacrt 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapien

<400> 40
tgggaactct gcayg 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapien

<400> 41
gaggctgagt gggrt 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapien

<400> 42
tggccacgcc ccayc 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapien

<400> 43
cttcctgagt ttcmg 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapien

<400> 44
cagatgaaga aackg 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapien

<400> 45
tatgcatttc atayg 15

<210> 46
<211> 15
<212> DNA
<213> Homo sapien

<400> 46
ttcctatcaa ctcrt 15

<210> 47
<211> 15
<212> DNA
<213> Homo sapien

<400> 47
cgagttgata ggarg 15

<210> 48
<211> 15
<212> DNA
<213> Homo sapien

<400> 48
atcctaccaa accyt 15

<210> 49
<211> 15
<212> DNA
<213> Homo sapien

<400> 49
gctgagggtg tgtmc 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapien

<400> 50
accagtgcgg atgka 15

<210> 51
<211> 15
<212> DNA
<213> Homo sapien

<400> 51
ctggctgtgt gaayg 15

<210> 52
<211> 15

<212> DNA
<213> Homo sapien

<400> 52
gtgacttggc cacrt 15

<210> 53
<211> 15
<212> DNA
<213> Homo sapien

<400> 53
acttcctaac aagrc 15

<210> 54
<211> 15
<212> DNA
<213> Homo sapien

<400> 54
gcagagaaag aggyc 15

<210> 55
<211> 15
<212> DNA
<213> Homo sapien

<400> 55
gctacttccg ggcrg 15

<210> 56
<211> 15
<212> DNA
<213> Homo sapien

<400> 56
gctgcagggc cccyg 15

<210> 57
<211> 15
<212> DNA
<213> Homo sapien

<400> 57
gggtgcaggt gggkc 15

<210> 58
<211> 15
<212> DNA
<213> Homo sapien

<400> 58
ctgccactgg gggmc 15

<210> 59
<211> 15
<212> DNA
<213> Homo sapien

<400> 59
ctttgtgtgg caamg 15

<210> 60
<211> 15
<212> DNA
<213> Homo sapien

<400> 60
tagccttcat gcckt 15

<210> 61
<211> 15
<212> DNA
<213> Homo sapien

<400> 61
gggctccaaa ggara 15

<210> 62
<211> 15
<212> DNA
<213> Homo sapien

<400> 62
aagtccctgc catyt 15

<210> 63
<211> 15
<212> DNA
<213> Homo sapien

<400> 63
cagggacttg accyg 15

<210> 64
<211> 15
<212> DNA
<213> Homo sapien

<400> 64
gagcagcagg ggcrg 15

<210> 65
<211> 15
<212> DNA
<213> Homo sapien

<400> 65
ctgcgcaggt aggsa 15

<210> 66
<211> 15
<212> DNA
<213> Homo sapien

<400> 66
gggaagtgcg cctsc 15

<210> 67
<211> 15
<212> DNA
<213> Homo sapien

<400> 67
cacagggatg ttgmg 15

<210> 68
<211> 15

<212> DNA
<213> Homo sapien

<400> 68
ctgtacttta agckc 15

<210> 69
<211> 15
<212> DNA
<213> Homo sapien

<400> 69
ctcaccatcc cccsc 15

<210> 70
<211> 15
<212> DNA
<213> Homo sapien

<400> 70
gttatatgag gagsg 15

<210> 71
<211> 15
<212> DNA
<213> Homo sapien

<400> 71
taactgctca cctkc 15

<210> 72

<400> 73
actgccctgt ttcyg 15

<210> 74
<211> 15
<212> DNA
<213> Homo sapien

<400> 74
gaggggcgag ggcrg 15

<210> 75
<211> 15
<212> DNA
<213> Homo sapien

<400> 75
cactctgtct cttra 15

<210> 76
<211> 15
<212> DNA
<213> Homo sapien

<400> 76
agagtttaga gatya 15

<210> 77
<211> 15
<212> DNA
<213> Homo sapien

<400> 77
tcagggcgtc cacyg 15

<210> 78
<211> 15
<212> DNA
<213> Homo sapien

<400> 78
acggtgtggg ggcrg 15

<210> 79
<211> 10
<212> DNA
<213> Homo sapien

<400> 79
ctttgtgcct 10

<210> 80
<211> 10
<212> DNA
<213> Homo sapien

<400> 80
caaccaggtc 10

<210> 81
<211> 10
<212> DNA
<213> Homo sapien

<400> 81
tccctatccc 10

<210> 82
<211> 10
<212> DNA
<213> Homo sapien

<400> 82
ggactgggaa 10

<210> 83
<211> 10
<212> DNA
<213> Homo sapien

<400> 83
acctgaggga 10

<210> 84
<211> 10

<212> DNA
<213> Homo sapien

<400> 84
ggtctcccgg 10

<210> 85
<211> 10
<212> DNA
<213> Homo sapien

<400> 85
agtcgcatcg 10

<210> 86
<211> 10
<212> DNA
<213> Homo sapien

<400> 86
gctcagagag 10

<210> 87
<211> 10
<212> DNA
<213> Homo sapien

<400> 87
ccaggctagt 10

<210> 88
<211> 10
<212> DNA
<213> Homo sapien

<400> 88
cataagtggg 10

<210> 89
<211> 10
<212> DNA
<213> Homo sapien

<400> 89
gcaggcgcac 10

<210> 90
<211> 10
<212> DNA
<213> Homo sapien

<400> 90
gaactctgca 10

<210> 91
<211> 10
<212> DNA
<213> Homo sapien

<400> 91
gctgagtggg 10

<210> 92
<211> 10
<212> DNA
<213> Homo sapien

<400> 92
ccacgcccca 10

<210> 93
<211> 10
<212> DNA
<213> Homo sapien

<400> 93
cctgagtttc 10

<210> 94
<211> 10
<212> DNA
<213> Homo sapien

<400> 94
atgaagaaac 10

<210> 95
<211> 10
<212> DNA
<213> Homo sapien

<400> 95
gcatttcata 10

<210> 96
<211> 10
<212> DNA
<213> Homo sapien

<400> 96
ctatcaactc 10

<210> 97
<211> 10
<212> DNA
<213> Homo sapien

<400> 97
gttgatagga 10

<210> 98
<211> 10
<212> DNA
<213> Homo sapien

<400> 98
ctaccaaacc 10

<210> 99
<211> 10
<212> DNA
<213> Homo sapien

<400> 99
gagggtgtgt 10

<210> 100
<211> 10

<212> DNA
<213> Homo sapien

<400> 100
agtgcggatg 10

<210> 101
<211> 10
<212> DNA
<213> Homo sapien

<400> 101
gctgtgtgaa 10

<210> 102
<211> 10
<212> DNA
<213> Homo sapien

<400> 102
acttggccac 10

<210> 103
<211> 10
<212> DNA
<213> Homo sapien

<400> 103
tcctaacaag 10

<210> 104
<211> 10
<212> DNA
<213> Homo sapien

<400> 104
gagaaagagg 10

<210> 105
<211> 10
<212> DNA
<213> Homo sapien

<400> 105
acttccgggc 10

<210> 106
<211> 10
<212> DNA
<213> Homo sapien

<400> 106
gcagggcccc 10

<210> 107
<211> 10
<212> DNA
<213> Homo sapien

<400> 107
tgcaggtggg 10

<210> 108
<211> 10
<212> DNA
<213> Homo sapien

<400> 108
ccactggggg 10

<210> 109
<211> 10
<212> DNA
<213> Homo sapien

<400> 109
tgtgtggcaa 10

<210> 110
<211> 10
<212> DNA
<213> Homo sapien

<400> 110
ccttcatgcc 10

<210> 111
<211> 10
<212> DNA
<213> Homo sapien

<400> 111
ctccaaagga 10

<210> 112
<211> 10
<212> DNA
<213> Homo sapien

<400> 112
tccctgccat 10

<210> 113
<211> 10
<212> DNA
<213> Homo sapien

<400> 113
ggacttgacc 10

<210> 114
<211> 10
<212> DNA
<213> Homo sapien

<400> 114
cagcaggggc 10

<210> 115
<211> 10
<212> DNA
<213> Homo sapien

<400> 115
cgcaggtagg 10

<210> 116
<211> 10

<212> DNA
<213> Homo sapien

<400> 116
aagtgcgcct 10

<210> 117
<211> 10
<212> DNA
<213> Homo sapien

<400> 117
agggatgttg 10

<210> 118
<211> 10
<212> DNA
<213> Homo sapien

<400> 118
tactttaagc 10

<210> 119
<211> 10
<212> DNA
<213> Homo sapien

<400> 119
accatccccc 10

<210> 120
<211> 10
<212> DNA
<213> Homo sapien

<400> 120
atatgaggag 10

<210> 121
<211> 10
<212> DNA
<213> Homo sapien

<400> 121
ctgctcacct 10

<210> 122
<211> 10
<212> DNA
<213> Homo sapien

<400> 122
tagctgcccg 10

<210> 123
<211> 10
<212> DNA
<213> Homo sapien

<400> 123
gccctgtttc 10

<210> 124
<211> 10
<212> DNA
<213> Homo sapien

<400> 124
gggcgagggc 10

<210> 125
<211> 10
<212> DNA
<213> Homo sapien

<400> 125
tctgtctctt 10

<210> 126
<211> 10
<212> DNA
<213> Homo sapien

<400> 126
gtttagagat 10

<210> 127
<211> 10
<212> DNA
<213> Homo sapien

<400> 127
gggcgtccac 10

<210> 128
<211> 10
<212> DNA
<213> Homo sapien

<400> 128
gtgtgggggc 10

<210> 129
<211> 18925
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (1128)
<223> PS1: Polymorphic base G or A

<220>
<221> allele
<222> (1632)
<223> PS2: Polymorphic base G or C

<220>
<221> allele
<222> (1760)
<223> PS3: Polymorphic base G or C

<220>
<221> allele
<222> (2047)
<223> PS4: Polymorphic base C or T

<220>
<221> allele
<222> (2146)
<223> PS5: Polymorphic base T or C

<220>
<221> allele
<222> (2279)
<223> PS6: Polymorphic base A or G

<220>
<221> allele
<222> (2634)
<223> PS7: Polymorphic base A or G

<220>
<221> allele
<222> (3840)
<223> PS8: Polymorphic base C or A

<220>
<221> allele
<222> (3884)
<223> PS9: Polymorphic base C or T

<220>
<221> allele
<222> (3897)
<223> PS10: Polymorphic base A or G

<220>
<221> allele
<222> (4055)
<223> PS11: Polymorphic base C or A

<220>
<221> allele
<222> (4308)
<223> PS12: Polymorphic base C or T

<220>
<221> allele
<222> (7462)
<223> PS13: Polymorphic base A or G

<220>
<221> allele
<222> (8710)
<223> PS14: Polymorphic base G or A

<220>
<221> allele

<222> (8955)
<223> PS15: Polymorphic base G or T

<220>
<221> allele
<222> (10907)
<223> PS16: Polymorphic base A or C

<220>
<221> allele
<222> (11040)
<223> PS17: Polymorphic base A or G

<220>
<221> allele
<222> (11058)
<223> PS18: Polymorphic base C or T

<220>
<221> allele
<222> (11424)
<223> PS19: Polymorphic base C or G

<220>
<221> allele
<222> (11486)
<223> PS20: Polymorphic base C or A

<220>
<221> allele
<222> (13913)
<223> PS21: Polymorphic base C or G

<220>
<221> allele
<222> (13936)
<223> PS22: Polymorphic base G or T

<220>
<221> allele
<222> (13982)
<223> PS23: Polymorphic base C or T

<220>
<221> allele
<222> (14559)
<223> PS24: Polymorphic base G or A

<220>
<221> allele
<222> (14714)
<223> PS25: Polymorphic base T or C

<400> 129
gatcacttga gggcaggagt tcaagaccag cctggccaaa atggcaaaac cctgtctcca 60
ctaaaaatac aaaaattagc caggtgtggt ggcacatggc tgtagtccca gctacttggg 120
aggctgaggc aggagaatca cttgaaccca ggaggcagag gttgcagtga gccaagatca 180
caccaccgca ctgcagcctg gtgacagagc acgactgtgt ctcaaaaaaa ttaattaatt 240
aattaaataa aaaaggaaat ggaactattt ttgtgaatct gtggattata tcagaaaaaa 300
aagacacaat ggggaaagtc ctaggcaaat caggatgagt tagtcatcct tcctagatga 360
gtgtttggtg ctaaatacat gctcagcaga catgattatt gcttcccctt tctttcgtcc 420
atttggcaac aaaaaggtgg caagcaccca ctctgtgccc tgtcctaggg tccgggaacc 480
ctgtaagcag tagatggagg tgggggtggg ggtgggggcg gggatgctgt tcagagcacc 540
ttgctccaag ggttcattaa aaaatccacc agtggaccgg gcgcggtggc tcatgccttt 600
aatcccagca ctttgggagg ccgaggcggg cggatcacaa ggtcaggaga tcgagaccat 660
cctggctaac acggtgaaac cccgtctcta ctaaaaatac aaaaaaaatt agccgggcgt 720
ggcagcgtgc gcctgtagcc ccagctgctg gggaggctga ggcaggagaa tggcgtgaac 780
ccgggaggcg gggcttgcag ggagccgaga tcgcgccact gcactccagc ctgggggaca 840
gagcgagact ccgtctcaaa aaataaaaat aaaaaaaata gaaaaaacaa tccaccagcc 900
acgataaatg gcagacctcc ttctgatttc agccggtgtg gtatgttcct gggctgacag 960
cacttgtcta gtcttgcttt cccaagtggg aaaggtctct gggaccttaa ggtccccagg 1020
tggtgacaca gagacaggta ggggggccca tagcaaagcc aggcaaggag gtcccgagat 1080
gattgtgggt ggcagggaaa gaaaaaatat tccttgactt tgtgcctrga cctggttgta 1140
ataaaggccc aagaggtagt tcctatcatc gtgcacattt cgctgaagga agaaactgag 1200
ggtcagtgac ccaagtgaag tgacttgccc aagatcatgc aggaagacat ggataattgt 1260
aatttgaacc aaggtcccag caaagtggga ttgttggggc tgagtgggcc ggctcctgca 1320
tttccttccc tctccctggg cttgggtctc ccacttgtcc agacagcggc cgggcttgtc 1380
acggggctct gtgcagcctt ttccactctc ccggctgcca gcgtcccgcc ccgtccctc 1440
ccagccccca agggaggagg ggagagctgc agagaggagg aggggtcggg gaggccggct 1500
ttataaaggc ggctggaaca accctgcccg ccagaccccg tcgcccggat cccctgagct 1560
gcccgccatc ccacgtgacc gcgccgcccc ccagctccac cgctgagtga gttggggcgc 1620
gtccctatcc csttcccagt ccctcgccga gccgcgtggt gcactgaagc cggcgtgggg 1680
aggagagccg atgaaggaga acgctaacat gggggctcca ggcagaatct ctaatgggag 1740
agatttagga cctgagggas ccgggagacc cgggagccca cggtctggtc ggccacctcc 1800
tctcctcccc gggcgcgagg cctatgagac cgcaagccac cctccaagaa tgcgcgtgca 1860
gtcgttgccg tggcaacgtg caagcctgca tgggcccacc gctaaaggaa aggggaggtt 1920
gttgggggtg tgggggctgg tcccctggag gtccaaactc atcacgcggc ggtagcacga 1980
gggaactggg aacgtcccct cccccgtgtc ccaaatcctc tgggagccag aacgcaagtc 2040
gcatcgyctc tctgagcctc agtttccaca tctgccatat ggaggcaaga ttattggttc 2100
caagttcccc gaagagttat ctggagggta acgatccagg ctagtyccca cttatggttg 2160
gacgtccaag gaatgggttt aagcggctca gataactcct gcgaccccac cccacccca 2220
ggcccgctcg ccatggccct cttcggggcc ctcttcctag cgctgctggc aggcgcacrt 2280
gcagagttcc caggctgcaa gatccgcgtc acctccaagg cgctggagct gggtaaggcg 2340
caggggcaga cagggacgga cgggagcgga cggggttggg gtcgccccaa cagtgggcac 2400
gggagctaag taagggtttc tctgctgctt cagtgaagca ggaggggctg cgctttctgg 2460
agcaagagct ggagactatc accattccgg acctgcgggg caaagaaggc cacttctact 2520

acaacatctc tgagtaagtg gggggcgggg cctcgcgatc cggggcgggg cctcggcggt 2580
ggaggcgggg cccgagaaat ctcagggtct cgggagactg gaggctgagt gggrtggggc 2640
gtggccaaat cgatgaggca agacttaaga aaccggatgg ggccgaaaag ataaaggcca 2700
agccagtcat cataggcggg accaggcaga tgggcgtggc cttgaaaatg tagagggggc 2760
caagaaggaa gcaggaccag tggatgtgag gcgtggtcga agaacaacag aaaggtggag 2820
tcaaggaact tcccaataga tgagagtgga gtctcgggca cggggtggag ttaatcggag 2880
gggtggagaa actgaaggct ctaagtatgg tttaggatga gggaggggga ataaggaatt 2940
gagatgtgag acggagctta ggtgatggag gcaaagtgaa cggactctat taaaagaccg 3000
ggtgtagcaa ccccagggtg tgggagggga gggggtgcaa gcaaacaaga ctttttttct 3060
ttttttttct gagagggagt cttgttctgt tgcccaggct agagtgcagt ggcgcgatct 3120
tggctcactg caacctccgc atcccgagtt caagcgattt tcctgcctca gcctcccgag 3180
tagctaggac tacaggcgcg tgccaccacg gtctgctaag ttttgtattt ttggtagaga 3240
cgggtttcac tatgttggcc agactggtct cgaactcctg acctcaagtg atccgccctc 3300
ctcggcctcc taaagtgctg ggattacagg catgagccac caccgcacct ggcaaacaaa 3360
caaacaagac ttttaagcga tggggtgaag ctcattttta acacctggat caggaagacc 3420
tcctctctca gccacccaag ttcaactttg agccacccaa gtctaagttg ggcacttgag 3480
ggacaatgga agaattcatt ccacccactc ccgcttcccc aggcctggac ttgaaagggg 3540
agcagacaaa tttcctgtcg ttgggggaag ttccctcttc ttggccctgg atctgaccct 3600
gaggcctcct gtagggtgaa ggtcacagag ctgcaactga catcttccga gctcgatttc 3660
cagccacagc aggagctgat gcttcaaatc accaatgcct ccttggggct gcgcttccgg 3720
agacagctgc tctactggtt cttgtaagga cccagcaccc tcaggggagt ggagagtggg 3780
tcaggagggc tggaggtgct gtggggctgt gggcagcccc cttagccttc ctgagtttcm 3840
gtttcttcat ctgtgtaaca gaaacaggca tatgcatttc ataygagttg ataggarggt 3900
ttggtaggat gtgtggaaag aggtgaaacc ttattaagag gtatggtgat gcatgggggc 3960
actgaagact caattgagct cctacctgga gtgaatatta acccccctgg cagctatgat 4020
gggggctaca tcaacgcctc agctgagggt gtgtmcatcc gcactggtct ggagctctcc 4080
cgggatcccg ctggacggat gaaagtgtcc aatgtctcct gccaggcctc tgtctccaga 4140
atgcacgcgg ccttcggggg aaccttcaag taagccccag ccccaacccc agctcaccct 4200
tttcgagccc cacagctcat ctgcctagcg caatgcaaag tgcataggcc tgggaatcag 4260
accaactggg tccaaagctc agctctgtca cttcctggct gtgtgaaygt ggccaagtca 4320
ctgaagcttg gtgcctcaat atcctcattt tataagatcc tattttataa gatcgttgtc 4380
aggattattt gtgttggtat atgtaaaaaa tatttagaac aaggctggct gtggtggctc 4440
atgcctgaaa tcccagcaat ttgggaggtc tgggcgggcg gatcacaagg ttaacagatc 4500
gagatcatcc tggccaacat ggtgaaaccc catctctact aaaaatacaa aaattagctg 4560
ggcgtggtgg cacgtgcctg tagtcccagc tacgtgggag gctgaggcag gagaatcgct 4620
tgaacccggg aggcggaggt tgcagtgagc tgagatcgtg ccactgcact ccagcctggt 4680
gacagagcaa gattccatct caaaaaaaaa aaaaaaaaaa aatttagaac agcacctggc 4740
atatgataaa tgtatattta agtgctggcc acttagataa gagtaactgt gaggactatc 4800
tgatatgcac atttggccat gtactggatt ctcaataatt ggtatctcta aataataata 4860
acaaaaataa taatttttt tttgagaggg agtctcactg tgtcacccag cctggagtgc 4920
agtggcgcca tctcggttta ctgcaacctc cacctcctgg gttcaaacga ttctcctgcc 4980
tcagcctccc aagtagctgg gactacaggc atgcaccacc atgcccaact actttttttg 5040
tatttttagt agagacaggg tttcgccata ttggccacgc tggtctcaaa ctcctgatct 5100
caggtggtct gtccgcctcg gcctaccaaa gtgctgggat tacaggcctg agcaactgtg 5160
cccggcctat ttttttattt ttgctgctga gaaggggata ttgttcttgt tctggaagag 5220
gatgggggag agaagatgat ggattataac ctggtgttgg tacactgtca gtattattta 5280
ttcatcaaac aaagtgtatg ccagtccctg aggatgcagg aaggagctca gagcctatct 5340
ggagcaaaaa acatttgata tcagggcctg gagagaagga ctaagaaaat gctttgggga 5400

ttcagaggag gggaatctgg gaagactgcc tggaagagga agcatctgag ctcagatggg 5460
aaaatacaga gtaggaatgc agagggcgga agggagggca tcagtaagcc gatggatgtg 5520
gggatgctca gagtgggttt gaggcaatgt gggtggattc atttgactga tgggaccaga 5580
gagtaggtca ggaggtcttg agaattagac tggagaggga atttgggcct ggttcttgga 5640
aaacctacgt ggaggagctg ttattcaggc tgaagctgag gactcctact gccactattt 5700
ccctagtcac tgatttctcc tggacctgga ctggggtgga ggcaggttct gggctcatcc 5760
ggcctctcct ctcctcacag gaaggtgtat gattttctct ccacgttcat cacctcaggg 5820
atgcgcttcc tcctcaacca gcaggtgtgg gcagcgacag gtcgcagggt ggcaagggtg 5880
ggcatgctct cactttgaga aggccctgac tctggctccc acctcgcaga tctgccctgt 5940
cctctaccac gcagggacgg tcctgctcaa ctccctcctg gacaccgtgc ctggtgagtt 6000
ggtggcgggt gagtctgggt gtgcagctgt catgcagcac ctcagagcag gccccttccg 6060
agccctgctg ttgaacagtc ctgggttcaa atgtggcccc tggagctgac ttgttgtgcg 6120
attttgggtg aagtgctcat ttctgtctgg gcaaagtgct tgtttcctca cttggctggc 6180
agtgcctcct cctggggttg ctgtgaggat tatctgagag aatagttgtc aaggtcctca 6240
atatgtgcat gctgatcact ggcatcccat ttgtcttcat cacctagcat ctaataggtg 6300
ctcaattaac atgaattccc tttttctttt tctttctttt ttttttttt ttgcacggag 6360
tcttacagtc tcactctgtt gccaggctgg agtccagtgg cgcgatctca gctcactgca 6420
atctccacct cccgggttca agcaattctc ctgcctcagc atctcgagta gctgggacta 6480
caggcgtgca ccaccatgcc caactagttt tttgtgtttt agtagagacg gggtttcacc 6540
atgttggccg ggatggtctc gatctcctga cctcctgatc tgcccacctt ggcctcccaa 6600
agtgctggga ttataggcat gagccactgt gcctggccat gaattccctt tttctaagcc 6660
atgacctctg tctactgtgg ggcaaaggta ttttcccatc acacctaaca cacaggcatg 6720
ggaagcgaag gaggaggtag gagaaaccaa gggaaaggaa cctcatgcca tctttttttt 6780
ttttttttt ttttgagatg gagtctcgct ctgtcaccca ggctggagtg cagtggtgca 6840
atcctggctc cctgcaacct ccacctccca ggatcaagcg attctcctgc ctcagcctcc 6900
tgagtagctg ggattacagg cgcccaccac catgcccaac taatttttgt atttttagta 6960
gagacggggt ttcaccgtgg tggaccaggc tggtctcgaa cacctgacct caggtgatcc 7020
gcccgccttg acctccctaa gtgctgggat tacaggtgtg agccaccatg cgcggccccc 7080
ccatgccatc ttgtttagtt attttttga aacctctata gtggtagtaa taataataac 7140
taacattaat gagcacctaa ctacacacca ggccctaagt gctttccaca taaaactaat 7200
ttgaccctca tgacaaccct gtgaaggatg tatcctcatt ttatgggtgg gcctctgagc 7260
ctggagaggt taagcacctt atcccacatc acacagccgt gagtgacttg agccccttcc 7320
tgcagtgcgc agttctgtgg acgagcttgt tggcattgac tattccctca tgaaggatcc 7380
tgtggcttcc accagcaacc tggacatgga cttccgggtg agctgcttgg gctggtgtat 7440
gacctctgac ttcctaacaa grcctctttc tctgctaagt tgcaagattt tacttcttaa 7500
agcacagctc tgatgaggcc actcccatta ctgagaatgc tatggctccc tgttactaca 7560
gaattaggtt cagcctcctg ggtctggcat ttgagcccct gttaatctta cctggcccca 7620
ccttttatat ctcatcctgt ttctttgcag ctgctctttc ccaaattgga ccctggcttc 7680
ccacctcagg gcctttgctc atgcacggcc tttccccaac tcccccgaag gagctggaca 7740
taaatttcta tacaaataat ctcatttcat tctctcacaa ccctgtgagt caggtactaa 7800
catcatccta attttaggaa tgtggagact tactcataga ggtgacgtca cttgcccaag 7860
gccacatagc cagtgtgtag ccactacctt tggtctccca gaggagtata gcagtttcag 7920
agtgcagact cgggagctag aatcagtcgg gtgcagtggc tcatgcctgt aatcccaata 7980
ctttgggagg cccaggcgag aggatcgctt gagcccagga gtttgagacc agcttgggca 8040
tcatagtgag acctccttct ctacaaagag aaaaagaatg cctgaagttc aaatcctagc 8100
tctgccactg actagctgca cggccctaag cataagtttt cagttggcca gtttcctaac 8160
atgtaaatgg aaataatcta gtaagaaaac acacgttgtt atgcagatta aagaagtcga 8220
tctatgtgaa gagcttagag tacttggcaa agaagcatga gtatgcacaa tatatgtaaa 8280

cagtagctgt tgttactgcc tctagctttc aaggcccact cagatgtcac ctcttccctg 8340
aaacctttcc tgaccctccc tcctccagtc tgacgggctc gccccctggt agcaggtgtc 8400
ctccttccaa tgccttttaa gagatgtgct tgtgaatcct ccaagaacag cagctgggtt 8460
cccaccatcc ccccacagtg cctagcacgg ggctgggaga cacctgctgt cagtccaggt 8520
gccaggccag tctgccttga ccctgactgt gaatgcccca ccccaggggg ccttcttccc 8580
cctgactgag aggaactgga gcctccccaa ccgggcagtg gagccccagc tgcaggagga 8640
agagcggatg gtgtatgtgg ccttctctga gttcttcttc gactctgcca tggagagcta 8700
cttccgggcr ggggccctgc agctgttgct ggtgggggac aaggtatgtc atggcctgtt 8760
tgtgggatgg gcaagagaag gtctgtgaca gagctcactc cctcactcct gattcccctg 8820
ttcaggtgcc ccacgacctg gacatgctgc tgagggccac ctactttggg agcattgtcc 8880
tgctggtgag tgctggcggg ggcagggata gggccacctg cacaccagtg agaccaagga 8940
ggggtgcagg tgggkccccc agtggcagcc acgcagacag ggccctggcc tcttgtctgt 9000
ggccaatccc cctcgccttg tcatggctgt tcctgtctgt gaagcggtct gctccttcct 9060
gctccttcct gccccatctc cttcgcagag ctgctgtgag gattagtgag aacccaagct 9120
ttgcaagtgt gagggcttta ataataattc acatatactt ttgtttattt gtgtattggg 9180
gaggagagta agagttacac attttattta ttttttttct tgagacaaga gtctcacttt 9240
gtcacccaag ttggaatgca gtggcacaat ctaggctcac tgcaacctct gcctcccagg 9300
gctcaagcca tcctcctgct tcagcctctt gagtagctgg aattacaggc ttgtgccacc 9360
acacccccgct aatttttata ttttgttgag acagagtttc attatgttgc tcaggctgtc 9420
cttgaactcc tgggctcaag tgatcaccca tttttggcct cccaaagtgc taggattaca 9480
ggtgtgagcc accgcaccta gcagaattac aaattttaaa agaaataatt cagtcattat 9540
catcatccaa tttatggcag ggcagggtct ctattgtcag actgcctggg cttcaagcct 9600
agttctgaaa ctgagtaact acataacctc tacaagttac ttaacctatc tgtgcctcaa 9660
tttccccatc tcaaaaatgg ggatatggct gggcatgatg gctcacgcct atactctgag 9720
aactttgaga ggtcaaggta ggaaaattgc ttgagcccag gagtttgaga ccagcctggg 9780
caacacagtg agacactgtc tctattgaaa aataaatttt tagtcagatc ttttttttte 9840
ggtgggggga caaagtctca ctctgtcacc caggctggag tgcagtggcg caatcttggc 9900
tcactgcaat tgaacctccc gggttcaagc aattctctgc catgacgact cagcctgaaa 9960
aataaatttg taaaaaataa caacaacaaa attggcagta ataatatcgg cctcacatgg 10020
ttattatgag gtttaaagga accaacacat atacagtgtt tagaatcatg tttggtgaag 10080
attaggcagt tgcatcttat agtattcctt cactgcctat agaagaaggc atagcagttt 10140
agagaaagaa gatatggcca ggcacggtgg ctcacgcctg taatcccaac actgtgggag 10200
gctgaagccg gcagatcact tgagcccagg agtttgagac cagcatgggc aacatggcca 10260
aaccccgtct ctacgaaaac acaaaagcaa aaaccaaaac aaaaaacaaa tacaaaaata 10320
agccaggcgt gatggcatga acctgtggtc tcagctgttc aggaggctga tgtggcagga 10380
taatttgagc tgaggaggtt gaggctgcag tgagccgtga tcatgccact gcactccagc 10440
ctgggcaaca gagaccctgt ctcaaaaaaa aaggagagag agagcaaaaa gatatggcaa 10500
ggggtttggg tgtggtggct catgcctgta atcccagcac tttgggaggc cgaggtgggc 10560
gaatcactag cctggccaat atggcgaaac cccgtttcta ttaaaaatac aaaacttagc 10620
cgggcatggc ggcaggtgcc tgtaatctca gctacttggg aggctgaggc gggagaatcg 10680
cctgaacctg ggagacagag ttgcagtgag ccgagattgt gccactgcac tccagcctgg 10740
gcaacaaagt gagactccat atcaaaaaca gaagacatgg caagggctga gatactaatg 10800
caaaaaggac atccacaaag agtctggatt ggagatagac ctgaatattg agatgtattt 10860
aggaagcaga aagagaagaa cacattcctg tttctttgtg tggcaamggc atgaaggcta 10920
gatgcagatg tgagagtttt ggaagccaga gctccctctg tatttgcaat atgagagtcc 10980
ccagaggaca gaagctgcac ctgccttcct tccccttctt ccctgggggc tccaaaggar 11040
atggcaggga cttgaccygc ccctgctgct catcccccag agcccagcag tgattgactc 11100
cccattgaag ctggagctgc gggtcctggc cccaccgcgc tgcaccatca agccctctgg 11160

caccaccatc tctgtcactg ctagcgtcac cattgccctg gtcccaccag accagcctga 11220
ggtccagctg tccagcatga ctatggtacg ggtcccagag tgggggctgc tggctgggg 11280
tcagggaaaa atgggctgtg ggttactgac cttctgggtc agtaacatcc tcctccccat 11340
gttcctgcag gacgcccgtc tcagcgccaa gatggctctc cgggggaagg ccctgcgcac 11400
gcagctggac ctgcgcaggt aggsaggcgc acttccctgc acctcagtga ttggaggaca 11460
atcacacctg tccacaggga tgttgmgctt aaagtacagt gatacctgta agccactgag 11520
agctggggtc tgtctttgtt aatgcatgcc aaagtgcttt ggtaaagtct gactcctctg 11580
cattctcaca tgtcccagag gcaggcaggg caggtattat gagagatgaa caaatgggtc 11640
cagcaatgtt aaataacatg cccacagtca cacagcaaag atcagaacat gggttttgac 11700
aaaaacctca gcacaacaca gctttgtgac cttgaggcca ttttatcttt tttaaaaatg 11760
ttattattct ttttttttt tttttgagaa ggagtttcgc tcttgttgcc caggctggag 11820
tgcaatagcg tggtcttgac tcactgccac ctctgcctcc cgggtttagg caattctgcc 11880
tcagcttcct gagtagctgg gattacaggt gtgcgccacc acgcccggct aatttttgca 11940
tttttagtag agacagggtt tcaccatatt ggtcaggctg gtctcaaact cctgacctca 12000
ggtgattcac ctgccccggc ctcccaaagt gctgggatta taggtatgag ccactgtgct 12060
tggccttatt ttattgtatt tatttatttt tgagacagag tctcactctg ttgcccaagc 12120
tggagtgcag tggtgcaatc atagctcact gcagccttga cctcctgggc tcaagtgatc 12180
gtccagcctc agcctcccga gtaactggga ctgcatgcat gcactaccat gcctgactat 12240
tttatttttt gtagagatgg ggtctcactt tgttgtccag gctggtctcc aattctcggg 12300
ctcaagtgat cgtcccacct tagcctccca aagtgctagg attacaggcg tgagccaccg 12360
cacccagccc attttgtctt tctaggcctc attttctcca tgtgtttaat acaggcaaaa 12420
gtagcatctg ccttctaggg ctgtgctatc caatatggta gatactagcc acatgtgact 12480
atttacattt aaattaatta taattaaata gtttttcagt tgcactaagc cacatttcat 12540
tagccacatg tggctagtga ctaccatagt agacggtgca gatacaaaac gtttcacaga 12600
aaatctgact ggacagtgct gttctaaggt taaaggagat aatgtaggtt aggaccctag 12660
gttctagagg cagcttctga cactttccag ctgtgtgact tcaggcaaat tacttaacct 12720
tgcagggcct cattttcctc atctataaat tagtgcaaat cctaatagtg cccactcctt 12780
gggcttgttt tgaggactaa accagttggt actgtatttt aaaacagttt actaatatta 12840
acttttaaa cacgttggct gggtgcagtg gctcacacct gtaatctcaa cactttggga 12900
gaccaaggcg ggtagatcgc ttgaggccag gagttcgata ccagcctgag ctacaaagtg 12960
agatcctgtc tctacaaaaa aatttaaaag ttagttggtc atagtgatgc acgcctgtgg 13020
tcccagctac ttggaagcct gaggtgggag gctgacttga gccctggagg tcaaagccac 13080
agtgagtaaa attacaccac tgccctccag cctgggcaac acagtgagac gctgtctcaa 13140
aaataaataa aaataaaaaa agaaaataaa tatgcttagt gtctgacagg aactaagcaa 13200
tgtgaagttt tgttacataa aataaataac acctgcaaag ctaccagcag aagtctggca 13260
cattagaagt gctcaataaa ggccgggcgt ggtggctcac gcctgtaatc ccagcacttt 13320
gggaggccga gacgggcgga tcacaaggtc aggaaatccg agaccatcct ggctaacacg 13380
gtgaaacccc gtctctacta aaaacacaaa aaattagctg ggcttggtgg cgggcaccta 13440
tagtcccagt tactcgggag gctgagccag gagaatggcg tgaacccggg agacggagca 13500
gtgagccgag atcgcgccac tgcactccag cgtgggcgac acagcgagac tctgtctcaa 13560
aaaaaaaaaa aaaaaaaaaa aaaaaagagg tgctcaataa ataaaaagta gcagtcttgg 13620
agaaaaggag acccaagtgg cagcaatgac tgccccgagg atacctggaa agttgtagca 13680
cttggaaagg tgcctaggtg ggcatctgag cctgcaactc agacacctct ggcaggtagg 13740
cctgggtggc ctgaggaagg ggagcttcca ggcttcctgc ccacagggct cctttctttc 13800
ccgcaggttc cgaatctatt ccaaccattc tgcactggag tcgctggctg tgagtgggga 13860
agggtgggtg ggctgggctg ggttcgtgag atcctgaccc tcaccatccc ccsctcctca 13920
tataactgct cacctkcggg cagctaaagg caaacccaac cctgactcac tgccctgttt 13980
cygccctcgc ccctcccagc tgatcccatt acaggcccct ctgaagacca tgctgcagat 14040

tggggtgatg cccatgctca atggtaaggc tggggtgtga ggatggagga agaaaggagg 14100
ggtgaactgg gcgggcccag actgagcggg gtgctcccac ccacagagcg gacctggcgt 14160
ggggtgcaga tcccactacc tgagggcatc aactttgtgc atgaggtggt gacgaaccat 14220
gcggtgagtg ggggcaggat ggggaaggag ggagggcctt cccctctctc tatagctccc 14280
cagtgccctc cattacctga ccccagctcc cccattacca gtcgcccctc ccattttaat 14340
cttagactcc taaaagccgt tgcacacact gtgacttctt ggagcactct tctcccatcc 14400
caatacctac taattgttca ggtcttagct cagaggtcac ctcctccaag aagccctcct 14460
tgacacccat agtctacact cctacagcct cattttaaca ccttgacatg gaattatcca 14520
ttagattgag ttcctgaaag acaaacactc tgtctcttra tctctaaact ctggcctcaa 14580
taactgaatg aataaatgag gctcccctct ccccaacaca gggattcctc accatcgggg 14640
ctgatctcca ctttgccaaa gggctgcgag aggtgattga gaagaaccgg cctgctgatg 14700
tcagggcgtc cacygccccc acaccgtcca cagcagctgt ctgagccctc aatccccaag 14760
ctggcagctg tcattcagga ccccaacccc tctcagcccc tcttttccca cattcatagc 14820
ctgtagtgcc ccctctaacc cccagtgcca cagagaagac gggatttgaa gctgtaccca 14880
atttaattcc ataatcaatc tatcaattac agtccgtcca ccacctccct gtgggctgtc 14940
ctgagctctg ttgggttcct gggatggaat cagtgcatca taaagggcat tctttaagca 15000
gagaaggggc caggccaccc cattcaggaa ctgctgcggg aataaagtgc taacttgccc 15060
ccaggctgtc tatgggagac cctgggccca gtctgggatg tacagggctc tgggaagggg 15120
gcagtcctgg cggcagaacc cggcctgcag gggcactttg cttagaagag gactctccta 15180
gcgggagagg ctgggagggg ctgcatcagg ccgtggagct ggttgctgtg gtcatcagta 15240
tggctgcttg ttcaggaagc gggagaacat ggtgaaggca gcgaggggct tgtcggtggg 15300
aaccatgtgg ccggcgccct gaggagcaat gttcgtgagc tcctgacccc accattccct 15360
cctccccata taactgctca ctcgggggca attccttcat cccaaaccct ttattcttcc 15420
cagaaccctc cccacccctc tccaaaaaaa cttgcccata caggggccag atggtgaccc 15480
atgacccagc ctaaaaggca gccagaggga aaggacgggt gggtcctgct cctttgcctc 15540
cggcccagtt atctctcagc aggcccagtc cctaccttga tcgtgagaaa ggcgatgtgg 15600
gagaactcct tcacgaagcc ggcaatctgc tccccgctgt ccccgtactt cactaaccag 15660
ggccggcgct gcacctccat ctgccccacc aggaaagaca tcagcctaca gcagctgcat 15720
ccttgctcac agctaccagc aagaccttag ggctgggaat tcctccacac ttgccctctg 15780
tgggccagag ccaggcagcc agctggccac tcccaggcat acccgctccc aatcctccac 15840
agcagcccct atcccagggc caggaatctc taccttacct tctggttgag ggaatccaca 15900
aaccactcat cccccatgaa attgcaggcc atgtctacat ctccattata taataggatc 15960
tggtatttct aaagcaggat ggggtaaaaa tgaggggtgt ggaacaagcc cagtccccag 16020
cccttcccta gttcaaggcc tacccctcag gaaattcaag gggccaagct agataacacg 16080
aaccagggaa ttttcatgtt ttctaacgac ttactgcatg tccagtattc tactaaatgt 16140
tttatctgtg aaggtagata tcattattat ccccatttta cagatggaga acctgaggct 16200
catatttaag aaatttgccc aagattatgt ggctagtaca gggacttgca cccaggccat 16260
gtgactccag agcccatgct attaaccact gtgctggaca gaacctagtc ctaacttggc 16320
cctgccagtg cttcctatat caccttgtgc aaggtctcct cccaccctgg acctctgctg 16380
gcccatttt ggcatgaatg aagctcttct caccccttaag actccgtggc tgatgcagac 16440
ccctaggcct cctatagcag gacagcatgg gtggggacaa acccaggttc aaattctgac 16500
cccaccaatc acaggaccct gggcttcgcc tttcggagtt tcagcttccc ttgtcctgac 16560
aggaagtctc aagggccccc agcttgctgt gctactctgg ggtcctaagg ctgctggtga 16620
tccagctgtg ctctccccac tcacctgtga gctaagcagc ttcagatact gggagttcat 16680
gcttcggtag agacggcggt actgtaagtt taccagaaag ctgccgggaa catacgagca 16740
ggccccagga agaccaggtc ggggcgaaaa gtagtcaaca gatcagacct tctctcctat 16800
ctccaagctc caacccagaa ggcagacagc tgggttccac tccccttttt gggggggggc 16860
cacctctcta ggactgactt ggatgttggg ctggatgggc tgggattcag acagcctcaa 16920

ctcaccacca tttagaaagg tacagggatg ctgggcgtgg tggctcacgc ctgtaatcct 16980
agcactttgg gagaccgagg tgagtggatc acgaggtcag gagttcgaga ccagcctgac 17040
caacatggtg aaaccccgtc tctactaaaa aaaaaaatac aaaaattagc caggcgtggt 17100
ggtgcgtgcc tgtagtccca gctactcaga aggctgaggc aggagaatca cttgaacccg 17160
ggaggcggag gttgcagtga gccaagatcg cagcactgca ctccagcctg agtgacagca 17220
agactccatc tcaaaaaaaa aaagaaaaag aaatgtacaa gggacgcatt ttggttgtta 17280
cgaggactgg gatgcatctg gcatttagtg tctaggggcc agcaggtgtg ggagtcacaa 17340
catacaacag ggtagagtgt gagcactgac caatcaaaat ggacctgagc agccctgctg 17400
ggcccagtgt taccctttta cttggggttg tgggggggag tggccacagg agactggaca 17460
tggaagttaa cagtttgtta ttgttgtttg agatggagtt tcacttttgt tgcccaggct 17520
ggagtgcagt agcatgatct cagctcactg caacctctgc ctcctgggtt caagcgattc 17580
tcctccctca gcctcctgaa tagctgggat tacaggtgcc caacagcaca cctggctaat 17640
tgttgtattt ttagtagaga tggggtttct tcatgttggc caggctggtc tcgaactcct 17700
gacctcaggt gattccccca cctcagcctc caaaagtgct aggattacag gcgtgaacca 17760
ctgcgcccag tcggaagtaa tagttattaa ccaatgtgat ggccgggtgt ggggaccctc 17820
gcctgtaatc ccagcacttt gggaggccaa ggagggagga ttgcttgaga ccaagagttt 17880
gagaccaacc tggacaacat agtgagaccc tgtctctaca aaaaaatttt ttaaaaatta 17940
gccaggaatg gtggcatgtg cctgtagacc cagctactca gtagactaag gtgggaggat 18000
cgcttgagcc taggaggttg aggctgtagt gaactgtgaa acaacaacca atgtggctac 18060
gtaatgaaat ggggctgggc ggggtggctc atgcctgtaa tcccagtact ttgggaggct 18120
gaaatgggca gatcacttga gccccagagt tcaaaaccag cctgagcaat gtggtaaaac 18180
accatctgta caaaaaatac aaaaacttgt tgggggtggt ggtgcgtctg tagtcccagc 18240
tactggggag gtgggaggat tgcttgagcc catgagtttg aggctgcagt gagctatgat 18300
cacaccactg cactccagcc taggtgacag agtgagacct tgtctcaaac aaaaaaattt 18360
taaatttaaa aaagataaaa gaatgaattg gctgactctc aatccactga ggggtcaggg 18420
atggatgtta aacatcctct acagctcatc tcattgagaa actctacaca gaagatccct 18480
gagaaacctg ccgatgccca caagccccag caacccacca ccccaagtca caggtggcca 18540
cggaacctca cttgcacatg tcccattgtg gcagctgctc cgggatgttg agggccttcc 18600
gcacgtacgg gttgttgagg taggtggaag cagctgttgt gttggtgcag ggggggtcca 18660
tgcgcacttt atcccctgag cgcagcagtg cctggaaggc acagacagtg ggtcagggcc 18720
tctgcttccc tgcccctccc cctgccccca cctcaccagg aggaagccca cgccctcgca 18780
cacctgatgc cacatccgct tgagtggcag gcgagtgaag atgttgccca aatcctggac 18840
cacaacagtg tccttctcat acctgcaatg tgaggtgggg gaatgggtgt ggagcctcca 18900
gatggggctg gggttggctc caggg 18925